CATIONS OF NON-PARAMETRIC DENSITY ESTIM

DISSERTATION

Ahmed Mohamed Mohamed Sultan
Lieutenant Colonel, Egyptian Air Force

AFIT/DS/ENC/90-1

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

1

AFIT/DS/ENC/90-1

APPLICATIONS OF NON-PARAMETRIC DENSITY ESTIMATION

DISSERTATION

Ahmed Mohamed Mohamed Sultan
Lieutenant Colonel, Egyptian Air Force

AFIT/DS/ENC/90-1

AFIT/DS/ENC/90-1

APPLICATIONS OF NON-PARAMETRIC DENSITY ESTIMATION

DISSERTATION

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Ahmed Mohamed Mohamed Sultan, B.S. , Diploma Degree , M.S.
Lieutenant Colonel, Egyptian Air Force

February 1990

AFIT/DS/ENC/90-1

APPLICATIONS OF NON-PARAMETRIC DENSITY ESTIMATION

Ahmed Mohamed Mohamed Sultan, B.S. , Diploma Degree , M.S.

Lieutenant Colonel, Egyptian Air Force

Approved:

Albert H Moore — Feb. 28, 1990
Chairman

Feb. 28, 1990

James R Simpson — Feb 28, 1990

Kenneth W Bauer — FEB 28, 1990

R. A. Calico Jr. — Feb. 28, 1990

R. A. Calico Jr.

Dean, School of Engineering

Preface

I would like first to thank Allah (God) whose help was more than necessary to finish my research and who facilitates all other means for me.

I am more than deeply indebted to the professor who has a meaning to me more than any words can express. This is Prof. A. H. Moore who had the idea of making nonparametric density estimation accessible by the practical application of the theoretical results of the subject. The professor who was always there whenever I have any kind of problems. I shall always be proud of being his student and I shall always remember his words to me. It was not just a professor student relationship, but I can simply say it was more than a father son relationship. I hope Prof. Moore is happy with my final effort and output, after a long preparation of my educational experience that started in 1981 and finished with this research. With his huge number of publications, research and technical expertise he was the best person to get help from during various phases of this research. I am also so thankful to Dr. Cain, Joseph P. to whom I am indebted for his interest in linear models when I chose this area as my master thesis research area. I am so thankful to my committee members Dr. Cain, Dr. Robinson, and Dr. Bauer for their friendly spirit, enthusiasm. and valuable comments and notes on the draft which helped me obtain a better final presentation of the different ideas of the research.

Finally and most importantly, to my wife, Azza, to my son Mohamed, my

lovely Egyptian-American daughter Dina, my beautiful twins Fatma and Maii, I like to express my love and appreciation for their continued support and help during my study at AFIT.

Ahmed Mohamed Mohamed Sultan

Table of Contents

List of Figures

Figure		Page

List of Tables

Table		Page

AFIT/DS/ENC/90-1

Abstract

The dissertation examines various methods of nonparametric density estimation, and nonparametric kernel estimation in more detail. The consequences of various kernel window width and their effect on the mean integrated square error are examined using Monte Carlo techniques.

The mean and the variance of nonparametric density estimator is derived for symmetric kernels with finite mean and finite variance. The results also treat kernels with varying window parameters.

The nonparametric kernel estimate was used to obtain new estimators for the three parameter Weibull distribution using distance estimation and the Cramer-von-Mises statistic. Comparison with maximum likelihood estimators using a Monte Carlo sample of size 1000 and various different parameters showed a significant improvement over the maximum likelihood estimators in the mean integrated square error between the estimated distribution and the true distribution.

Several new goodness of fit tests are proposed using the nonparametric kernel estimator and the Cramer-von-Mises and the Anderson Darling statistics. Extensive Monte Carlo experiments were performed to obtain the critical values for the test and to study the power of the tests against eight alternative distributions. The tests using the Anderson Darling statistic showed greater power against almost all alternative distributions studied than the K.S. test.

A new nonparametric kernel estimator was introduced by varying the window width in each tail portion of the sample. The method permitted different window width in each tail portion and in the center portion of the sample. The method uses separately the sample percentile ratios as a measure of each tail length. The kernel parameter for the tail sample values is chosen using sample percentile ratios for that tail. The nonparametric kernel estimator results in comparable mean integrated errors with the estimators developed earlier.

APPLICATIONS OF NON-PARAMETRIC DENSITY ESTIMATION

I. Introduction

The idea of using nonparametric density estimation is a rich research topic. both in estimation techniques and in applications. Two previous dissertations under the supervision of Prof. A. H. Moore studied density estimators with applications (Sweeder, 1982 and Fuchs, 1984) .

A continuation of the previous research, with the idea of exploring some new applications of the nonparametric density estimation, using different nonparametric density estimators, is the goal for this research.

This dissertation is divided into six main parts (chapter II-VII). The first part surveys some of the known nonparametric density estimation methods with the aim of looking at the different results and deciding which of these methods meets the need for a nonparametric density estimation technique with the least number of parameters and the most established theoretical results. Among these methods are the orthogonal series method, the penalty functions method. the delta sequence method, and the nearest neighbor method. This part is briefly concluded with a descriptive comparison from the literature of these methods.

Next, the kernel method which has (1) only one parameter, (2) the best understood properties, (3) the invariance property with respect to both location and scale. and (4) is computationally effective, is discussed in chapter III. Since the choice of the kernel is not as crucial as the choice of the parameter (window width h) in the kernel method, the Gaussian kernel is chosen which has an infinite support and solves the problem of finding the estimated density support when using a kernel with a finite support. In this chapter it is also shown that for the kernel method, the mean of the nonparametric density is the sample mean, and the variance of the nonparametric density is the sample variance plus the kernel variance. Since for certain applications the invariance of the density estimator is required, the invariance property for the kernel estimator is also shown. A suggested h is then introduced, based on the approximate optimal choice of the window width. A Monte Carlo experiment is designed with this proposed choice of h. The mean integrated square error (MISE) is used as a measure for the closeness of the true density to the estimated one. The results from different distributions are reported for sample sizes 10(10)60.

In chapter IV a numerical optimal choice of h is derived in the form of a constant multiple of the unbiased estimator of the standard deviation divided by the fifth root of the sample size. The different values of the constant of multiplication together with the corresponding h and MISE are reported for various distributions and a given sample size.

Chapter V and VI consider parameter estimation for the three parameter

Weibull distribution. In chapter V the log-likelihood equations are solved numerically using the hybrid method. Chapter VI considers the use of the minimum distance estimation technique to estimate the parameters of the three parameter Weibull distribution using Cramer von Mises statistic as a measure for the closeness of the density function with parameters obtained by the maximum likelihood method and the density function with parameters obtained by the new minimum distance estimation method. Results from a Monte Carlo of size 1000 are reported for both methods. The results demonstrate an improvement of the new estimation technique over the maximum likelihood technique.

In chapter VII a new modified goodness of fit technique for normality is introduced. The critical values for the test are generated. The power of the test for various alternative distributions is computed.

Chapter VIII introduces an adaptive density estimation based on the choice of different h for each tail of the distribution. The sample percentile ratios are used as a criterion for the choice of h in the tail values of the sample.

II. Survey Of Some Nonparametric Density Estimation Methods

Introduction

A large number of methods of nonparametric density estimation have been proposed. These methods have the common goal of estimating a density function when a set of data is given. A few types of estimates were first proposed in Fix and Hodges (1951). Although the nonparametric estimates involve some parameters, they are still considered nonparametric in the sense of relaxing the assumptions about the distribution of the observed data.

Many different methods of density estimation have been introduced and studied for a long time. Monte Carlo comparisons have been done for various nonparametric estimators. A discussion of some properties and basic results of the following nonparametric density estimation methods: orthogonal series method, the penalty function method, the delta sequence method, and the nearest neighborhood method will be surveyed and discussed in this chapter. The survey and discussion in this chapter follow essentially the discussion by Paraska (1983:27-173). The kernel method will be treated and studied in a separate chapter with a Monte Carlo experiment for different sample sizes from various distributions since it is the method that will be used for the different applications in the dissertation together with the reasons for choosing this method.

In the orthogonal series method, the density function is expressed in terms of its orthogonal series expansion and by estimating the coefficients in the orthogonal expansion the estimate of the density can be found.

In the method of penalty functions, an estimator is obtained through optimizing (maximizing) the likelihood function of the sample over densities such that the likelihood function has a finite maximum when the underlying density belongs to the class of density functions.

The delta sequence method is in fact a generalization of other methods such as Fourier inversion and the kernel method.

The nearest neighborhood method is based on fixing a constant r and choosing the r^{th} ordered distance of all the observations from a given point, then using this distance as a smoothing parameter.

The method of kernels is a widely used method in applications with the best understood properties. It came from the idea of the naive estimator, which is an evolution of the histogram as will be discussed later in chapter III.

The performance of the different methods have been reported in the literature and the methods studied, hence in the next part a summary of each of these methods is presented.

The Method Of Orthogonal Series

The idea of this method is to express the density function in terms of its orthogonal series expansion and to find the estimate of the underlying density through estimating the coefficients in the orthogonal expansion.

This method was first introduced by Cencov in a 1962 paper. To show the conditions under which it is possible to expand a function f in terms of a set of a complete orthonormal basis, let us assume that $\mathcal{X}$ is a space, $\mathcal{M}$-is a σ- algebra of subsets of$\mathcal{X}$ i.e

$$\emptyset \in \mathcal{M} \ , \tag{1}$$

$$E \in \mathcal{M} \Longrightarrow \mathcal{X} \backslash \mathrm{E} \in \mathcal{M} \ , \tag{2}$$

$$E_j \in \mathcal{M} \Longrightarrow \cup_\mathrm{j} \mathrm{E_j} \in \mathcal{M} \tag{3}$$

where $\cup_\mathrm{j}$ is the union over j

Hence$(\mathcal{X}, \mathcal{M})$ will be a measurable space.Let $\mu : \mathcal{M} \rightarrow [0, \infty]$ be a measure and the norm space $L^2(\mu)$ is separable. i.e

$$\mu(\emptyset) = 0 \ , \tag{4}$$

$$\{E_j\}_{j \geq 1} \subset \mathcal{M} \Longrightarrow \mu(\cup \mathrm{E_j}) = \sum \mu(\mathrm{E_j}) \tag{5}$$

where E_j are disjoint.

If $\mathcal{P}$ represents the family of probability measures on $(\mathcal{X}, \mathcal{M})$ such that the Radon - Nikodym derivative $dp/d\mu \in L^2(\mu)\ \forall p \in \mathcal{P}$, let $\mathcal{B} = \{\mathrm{b_i}, \mathrm{i} \geq 1\}$ be a complete orthonormal basis for $L^2(\mu)$. Since $\mathcal{B}$ is complete, then $f = dp/d\mu$ can be written as:

$$f(x) = \sum_{i=1}^{\infty} a_i b_i(x) \tag{6}$$

where

$$\begin{aligned} a_i &= \int_{\mathcal{X}} f(x) b_i(x) d\mu(x) \\ &= E_f b_i(x) \end{aligned} \tag{7}$$

which will correspondingly introduce the orthogonal series estimator of f based on a random sample of size n to be defined as:

$$\hat{f}(x) = \sum_{i=1}^{L_n} \left[1/n \sum_{r=1}^{n} b_i^0(x_r) \right] b_i(x) \tag{8}$$

where b_i^0 is a fixed version of $\mathcal{B}$, and the number of terms in the expansion $L_n \longrightarrow \infty$ as n $\longrightarrow \infty$ and where a_i is replaced by its estimator

$$\begin{aligned} \hat{a}_i &= \hat{E}_f[b_i(x)] \\ &= 1/n \sum_{r=1}^{n} b_i^0(x_r) \end{aligned} \tag{9}$$

The properties of $\hat{f}$ are studied in Bosq (1970) and can be briefly summarized in the following points:-

1. $MISE \longrightarrow 0 \iff \lim_{n\to\infty} \int_{\mathcal{X}} 1/n \left[\sum_{i=0}^{L_n} b_i^2(x)\right] f(x)\, d\mu(x) = 0$

where MISE is the mean integrated square error defined as:

$$MISE = E\int [\hat{f}(x) - f(x)]^2\, dx \tag{10}$$

2. If

(i) f is continuous.

(ii) $\mathcal{B}^* = \{b_i^*, i \geq 1\}$ are continuous and

$$M_n = \sup_{1\leq i\leq L_n} \sup_{x\in\mathcal{X}} |b_i^*(x)| < \infty, n \geq 1$$

(iii) $\sum_{i=1}^{\infty} a_i b_i^*(x) \overset{\text{uniformly}}{\longrightarrow} f(x)$, and

(iv) $\lim_{n\to\infty} M_n^4(L_n^2/n) = 0$

then

$$\lim_{n\to\infty} \sup_x E[|\hat{f}(x) - f(x)|^2] = 0 \tag{11}$$

3. If

(i) $\mathcal{B}^*$ is uniformly bounded

(ii) $\sum_{i=1}^{\infty} a_i b_i^*(x) \overset{\text{uniformly}}{\longrightarrow} f(x)$

(iii) $\exists$ m > 0 $\ni$:

$\int_{\mathcal{X}} \mathrm{b}_i^2(\mathrm{x})\,\mathrm{f}(\mathrm{x})\,\mathrm{d}\mu(\mathrm{x}) \geq \mathrm{m}$, i ≥ 1

(iv) $\lim_{n\to\infty} L_n = \infty$ and

(v) $\sum_{n=1}^{\infty} L_n\, exp(-\lambda n / L_n^2) < \infty, \forall \lambda > 0$

then

$$d^* = \sup_x |\hat{f}(x) - f(x)| \longrightarrow 0 \; as\, n \to \infty \tag{12}$$

4. If

(i) b_i is of bounded variation for all $i \geq 1$

(ii) $\sum_{i=1}^{\infty} a_i b_i(x) \overset{\text{uniformly}}{\longrightarrow}$ f(x)

(iii) $\lim_{n\to\infty} L_n = +\infty$

(iv) $\sum_{n=1}^{\infty} exp(-\alpha n \,/ M_{L_n}^2\, V_{L_n}^2) < \infty \,, \forall \alpha > 0$

with

$M_n = \sup_{1 \leq i \leq n} \sup_{x \in \mathcal{X}} |\mathrm{b_i(x)}|$,

$V_n = \sum_{i=1}^{n} \int_{-\infty}^{+\infty} |db_i|$

then

$$\sup_{x \in \mathcal{X}} |\hat{f}(x) - f(x)| \xrightarrow{a.s} 0 \text{ as } n \longrightarrow \infty \tag{13}$$

The necessary and sufficient conditions for the convergence of the density estimator using the method of orthogonal functions are given by Bosq and Bleuez (1976). Finally the advantages and disadvantages of the method will be stated in the discussion section at the end of this chapter.

The Method Of Penalty Functions

This method is characterized by applying the known methodology of estimation ,the maximum likelihood method , originally introduced by R. A. Fisher , which is considered as a universal method for optimal estimation.

The problem statement in this case is to find an estimate of the underlying density function from which a sample of size n was drawn such that the likelihood function is maximized.This is mathematically formulated as:

$$Max\ l\,(f|x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i) \tag{14}$$

where $x_1, ..., x_n$ are i.i.d random variables with a common unknown density f and l is the likelihood function of the sample .This likelihood function does not have a finite maximum when f belongs to the class of density functions $\mathcal{F}$. This makes it necessary to set restrictions on $\mathcal{F}$ to avoid that infinite solution.

An approach for using the maximum likelihood principle is by penalizing those functions giving an infinite solution. This infinite solution will essentially happen if $\mathcal{F}$ is a sequence of functions that converges pointwise to a Dirac - delta function. This means that the penalization would represent a way of deciding between smoothness and goodness of fit.

Now, define a penalty function $\mathcal{P} : \mathcal{F} \mapsto \mathcal{R}$ as a real-valued functional over $\mathcal{F}$; also define $L(f) = log\, l = \sum_{i=1}^{n} \log f(x_i)$ as the likelihood function and define

$$LP : f \mapsto L - \alpha P \;, \alpha > 0 \tag{15}$$

as the logarithm of the penalized likelihood function.

Hence, the problem will be to find a measurable function $\hat{f} : \mathcal{R}^n \mapsto \mathcal{F} \ni LP$ is maximized . This F is called *the maximum penalized likelihood estimator of* f.

A suggested penalty function (Good and Gaskin 1971) has the form:

$$P(f) = \int_{-\infty}^{+\infty} \left[f'(x)/f(x)\right]^2 dx \tag{16}$$

and the problem is formulated as:

$$MaxLP(f) = L(f) - \alpha P(f) \tag{17}$$

subject to :

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \tag{18}$$

$$f(x) \geq 0 \tag{19}$$

$$f(x_i) > 0 \quad ,\forall i = 1,...,n \tag{20}$$

and

$$P(f) < \infty \tag{21}$$

To avoid the non-negativity constraint Good and Gaskin used the substitution f=g^2 which transforms the problem to:

$$MaxLP(f) = 2\sum_{i=1}^{n} log|\,g\,|(x_i) - 4\alpha \int_{-\infty}^{+\infty} g^{'^2}(x)\,dx \tag{22}$$

subject to

$$\int_{-\infty}^{+\infty} g^2(x)\,dx = 1 \tag{23}$$

$$\int_{-\infty}^{+\infty} g^{'^2}(x)\,dx < \infty \tag{24}$$

and

$$|\,g\,|(x_i) > 0 \;\; \forall\, i = 1,..,n \tag{25}$$

The estimator obtained by this method is a spline function with double exponential splines and knots at the sample points.

An optimal solution for this problem which is twice differentiable with the same sign for all x was derived by Ghorai (1977).

The Method Of Delta Sequence

This method generalizes other different methods such as Fourier inversion method, Kernel method, Histograms and others. To define a delta sequence let $\mathbf{\Phi}$ be an element of the class of continuous functions with continuous derivatives of all orders i.e $\mathbf{\Phi} \in \mathcal{C}^{\infty}$ with support I=(a,b), a,b $\in \mathcal{R}$,for every $x \in$ I. $\Delta = \left\{\delta_i(x,t)\right\}$ is a delta sequence on I if:

$\delta_i : I \longrightarrow I \ni \delta_i$ is bounded measurable $\forall$ i= 1,... and

$$\lim_{i \to \infty} \int_I \delta_i(x,t)\mathbf{\Phi}(t)\,dt = \mathbf{\Phi}(x) \tag{26}$$

An estimator based on that method and an i.i.d sample $\mathbf{x_1}, \ldots, \mathbf{x_n}$ from f(x) would have the form:

$$\hat{f}(x) = \frac{1}{n}\sum_{j=1}^{n}\delta(x, X_j) \tag{27}$$

which gives a sequence of estimators when using the sequence Δ. This estimator can give other kinds of estimators like the ones mentioned above by a proper choice of the delta sequence. The necessary and sufficient conditions for the asymptotic

unbiasedness for some delta sequence based estimators are given by Walter and Blum (1976), while the asymptotic normality of such estimators is studied by Watson and Leadbetter (1964).

The Nearest Neighbor Method

This method is based on the choice of a fixed constant r, and through ordering the distance of each of the n observations from a given point one will be able to pick the r^{th} ordered distance. The mathematical formulation for this method comes from the idea that the number of observations in an interval of width $2w_r$ centered at $\hat{x}$ is exactly r-1.This implies that:

$$r - 1 = 2w_r n f(t) \tag{28}$$

which means that :

$$\hat{f}(t) = \frac{r-1}{2w_r n} \tag{29}$$

which gives the estimate of $\hat{f}(t)$ based on the r^{th} nearest neighbor. The method does not give an estimated density that integrates to one. The estimator for this method has discontinuous derivative at the points $\frac{(x_i + x_{i+r})}{2}$

Discussion Of Different Methods

This section surveyed some results about the kernel method, orthogonal series method, the method of penalty functions and the nearest neighbor method.

It is well known that in order to have a successful use of the non-parametric density estimation techniques there should be a sufficient amount of data and a reasonable information about the form of the underlying density function. A Monte Carlo study to compare density estimators of both the kernel method and the method of orthogonal series for specific distributions (*normal, uniform ,... etc*) is performed by Kumar and Markmann (1975).

In kernel estimation one must choose the kernel and the window width. The choice of the kernel does not significantly affect the efficiency of the estimator, however the window choice varies both *the bias* and *the variance* of the estimator of $f(x)$ at each value of x. Since the underlying density is not known, this means that there will be no guarantee that the choice of the window is the optimal one. However the kernel method gives an estimator which is a density when choosing the kernel as a density, besides being computationally efficient.

In orthogonal series estimation one has to choose *a basis* and some *cut off sequence.* The choice of basis will affect the mean integrated square error. The disadvantage of this method is that the basis is *arbitrarily* chosen independent of the given data, and the it *does not* give estimates which are densities. Furthermore, the estimators *could* be negative. However it is more efficient *computationally* than the kernel method since few terms give a sufficiently accurate estimator. A cosine-based estimate has been suggested by Anderson (1969) to have good characteristics.

In the method of penalty function, there is some complexity involved in the

calculations of the estimator, however using a discrete maximum penalized likelihood it becomes less complex. An advantage of the method is the insurance of the non-negativity of the estimator since the penalty function is a function of the logarithm of the density.

The nearest neighbor method was developed to find a commputationally fast technique for estimating the density. Contrary to the kernel method, this method over-smooths the distribution tails. Also, the estimates in this case are not every where differentiable and in general it does yield an estimator that integrates to unity.

However, after examining the methods discussed above in detail, the kernel method is chosen to be used for the applications studied in this reseach due to its following properties:

(1) It is Scale and location invariant if one chooses the parameter to be scale invariant.

(2) It gives a proper density function when the kernel is a density function.

(3) It does not give a negative estimator.

(4) It has only one parameter.

(5) It directly picks the support.

(6) It is fast in computations.

III. Monte Carlo Comparison For Some Distributions Using The Kernel Method

Introduction

The histogram as a basic model for density estimation, and the naive estimator are introduced in this chapter. The kernel method is then surveyed as being a natural evolution of the naive estimator. Some basic properties and results for the kernel estimator, together with some different kernels are introduced. The mean and variance of the kernel density are derived. The invariance property for the kernel method is shown. Finally a Monte Carlo experiment is designed to examine the behavior of a set of different distributions under a proposed choice for the window width of the kernel estimator. The experiment uses different distributions with the mean integrated squared error as the criteria for the comparison.

The Histogram

The histogram, if it is constructed so it integrates to one, is simply an estimate of the p.d.f as a function which varies based on a predetermined division of the support of the estimator. It also is expressed as a function of the number of observations from a sample of size n $(X_1, X_2, \cdots, X_n)$; in each of the subdivisions or

mesh of the support in the following way:

$$\hat{f}(x) = \frac{1}{nh}(\# of X_i in\, the\, same\, bin\, as\, x) \tag{30}$$

where h represents the width of each mesh or bin and known as bin width.

The bin width h can be allowed to vary in which case the form for the estimator will be:

$$\hat{f}(x) = \frac{1}{n}\left(\frac{\# of X_i\, in\, the\, same\, bin\, as\, x}{width\, of\, bin\, containing\, x}\right) \tag{31}$$

The basic properties for this estimator can be summarized as:

- Simple and easy way of data representation.

- It depends on the choice of origin and bin width.

- In bivariate and trivariate samples, it depends on the grid direction of the cells (besides origin and bin width)

The Naive Estimator

Since f(x) can be expressed as the limit of the rate of change of F(x) then

$$\begin{aligned} f(x) &= \lim_{\Delta t \downarrow 0} \frac{F(x+\Delta t) - F(x)}{\Delta t} \\ &= \lim_{\Delta t \downarrow 0} \frac{F(x+\Delta t) - F(x-\Delta t)}{2\Delta t} \end{aligned} \tag{32}$$

Hence, it is reasonable to estimate f(x) by $\hat{f}(x)$ as

$$\hat{f}(x) = \frac{F_n(x + \Delta x) - F_n(x - \Delta x)}{2\Delta x} \tag{33}$$

where

$$F_n(x) = \frac{no.\ of\ X_i's\ \leq x}{n} \tag{34}$$

now, using the conventional notation for the bin width as h_n instead of Δx which varies with the sample size n then

$$\hat{f}(x) = \frac{1}{n\, h_n}[no.\ of\ X_i's\ \in\ (x - h_n, x + h_n)\,/\,2] \tag{35}$$

where $h_n \longrightarrow 0$ as $n \longrightarrow \infty$

This estimator is known as the naive estimator. The naive estimator can be considered as a histogram with each observation as a center of a sampling interval.

This method gives a discontinuous estimator with jumps at $X_i \pm h_n$ and with zero derivatives everywhere else.

The Kernel Method

In this section, a more detailed discussion of the kernel estimators with their properties is considered. The naive estimator involves the idea of looking for a function through which one is able to obtain a measure for the count of the number of X_i's in the interval (x - h_n , x + h_n). Such function is known as kernel function

K(.) satisfying the regularity conditions:-

(i) sup K(x) $\leq M < \infty$, | x | K(x)$\rightarrow 0$ *as* $|x| \rightarrow \infty$.

(ii) K(x) is symmetric, $\int_{-\infty}^{+\infty} x^2 K(x)dx < \infty$.

(iii) K(x) has an absolutely integrable characteristic function.

and the estimator suggested in this case will have the form:-

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \tag{36}$$

where $h_n \rightarrow 0$ as n $\rightarrow \infty$.

The previous discussion gives a brief introduction to the concept. This concept can be summarized, in the case of the univariate spaces with continuous variables, as placing a kernel at each point of the design sample $\{X_1, ..., X_n\}$. Averaging the contributions of the different kernels at all points of the support results in the kernel estimator.

In spite of the fact that the kernel estimator resolves the major difficulties with the histograms. Such difficulties are the fixed cell structure, the discontinuities at cell boundaries, the lack of tails, and the exponential increase of the number of cells with the increase of the number of variables. The kernel estimator has the problem of the choice of the proper h_n. It is obvious that for a fixed n, a large h_n gives a very smooth estimate, and a small one gives an irregular estimate. It is noted that as $h_n \rightarrow 0$ the nonparametric density converges to a series of spikes at each of the

observations. This means that a difficulty corresponding to the choice of the cell size in the histogram will remain.

The mathematical properties for the univariate kernel estimators are well known. These include the bias and the asymptotic results.

The asymptotic properties of such an estimator are investigated by Parzen (1962) . The necessary and sufficient conditions for the uniform consistency with probability one for kernel estimators are studied by Nadaraja (1965) and Schuster (1970). Based on their study for the properties of the kernel estimator the following theorem holds:-

1. For a kernel function K(.) which is of bounded variation and $\sum_{j=1}^{\infty} exp(-\gamma j {h_n}^2)$ converges $\forall \gamma > 0$.

Then

$\delta = sup_x |\hat{f}(x) - f(x)| \longrightarrow 0$ with probability 1 as n $\rightarrow \infty \iff$ f is uniformly continuous.

Now, several results are introduced on the consistency of kernel estimator. First define:

$J = \int |\hat{f} - f|$

then the following results hold:

1. If the kernel is Borel measurable function on $\mathcal{R}^n \ni : K \geq o \ , \int K = 1$

then

(i) J $\xrightarrow{in\ probability}$ 0 as n $\longrightarrow \infty$ for some f.

(ii) J $\xrightarrow{in\ probability}$ 0 as n $\longrightarrow \infty$, $\forall$ f.

(iii) J $\xrightarrow{almost\ surely}$ 0 as n $\longrightarrow \infty$, $\forall$ f.

(iv) J $\xrightarrow{exponentially}$ 0 as n $\longrightarrow \infty$, $\forall$ f.

where the exponential convergence means : given $\epsilon > 0$, $\exists\ r$, $n_0 > 0 \ni$:

P(J $\geq \epsilon$) $\leq exp(-rn)$, $n \geq n_0$.

(v) $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} n(h_n)^m = \infty$.

2. For any density f on $\mathcal{R}^m$

; K is an absolutely integrable function $\ni$: $\int$ K = 1

; $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} n\,(h_n)^m = \infty$

then

J $\xrightarrow{exponentially}$ 0 as n $\to \infty$, $\forall$ f.

3. If K, f are densities on $\mathcal{R}^m$; J $\xrightarrow{in\ probability}$ 0 as n $\longrightarrow \infty$

then

$\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} n\,(h_n)^m = \infty$.

4. $\sup_x |\hat{f}(x) - E[\hat{f}(x)]| \xrightarrow{a.s} 0$ as n $\longrightarrow \infty$, $\forall$ distributions F.

5. Let B(x) = E[$\hat{f}$(x)] - f(x) be the bias of the estimator, and

$$m_2 = \int_{-\infty}^{+\infty} x^2\, \mathrm{K}(x)\, dx$$

where K(x) satisfies :

(i) $\sup_x K(x) \leq M < \infty$; $|x|K(x) \longrightarrow C$ as $|x| \longrightarrow \infty$.

(ii) $K(x) = K(-x)$, $x \in \mathcal{R}$; $\int_{-\infty}^{+\infty} x^2 K(x)dx < \infty$.

If f is a bounded density function and if $f''(x)$ exists, then

$$\lim_{n\to\infty} \frac{1}{(h_n)^2} B(x) = -\frac{1}{2} m_2 f''(x)$$

The choice of the smoothing parameter h is more crucial than the choice of the kernel itself. The approximate MISE as a function of h is given by :

$$\frac{1}{nh}\int K^2(t)dt + m_2^2\frac{1}{4}\int\{f''(x)h^2\}^2\,dx \tag{37}$$

which upon differentiation w.r.t h and equating to zero will give the optimal h_{opt}.

$$h_{opt} = m_2^{-2/5}\left\{\int K^2(t)dt\right\}^{1/5}\left\{\int f''(x)^2\,dx\right\}^{-1/5} n^{-1/5} \tag{38}$$

The h value gets bigger as the second derivative of f(x) gets smaller and consequently this gives a smoother estimator and a smaller approximate MISE.

An approach for the kernel choice that uses calculus of variation to derive a kernel that optimizes the approximate MISE gives a kernel with an efficiency 1 which is known as Epanechnikov kernel. This kernel is given as:

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}) & \text{if } -\sqrt{5} \le x \le \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$$

Defining the efficiency of a kernel as the ratio of its MISE relative to that of Epanechnikov, the relative efficiency of different kernels are given in Table 1.

Table 1. Different Kernels with their Efficiency

kernel	$K(x)$	*Efficiency*
Epanechnikov	$K(x) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}) & \text{if } -\sqrt{5} \le x \le \sqrt{5} \\ 0 & \text{otherwise} \end{cases}$	1.00
Boxed	$K(x) = \begin{cases} \frac{1}{2} & \text{if } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$	0.9295
Biweight	$K(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2 & \text{if } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$	0.9939
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} exp - (\frac{x^2}{2})$	0.9512

The idea of chosing a smoothing parameter value that will subjectively agree with a priori information about the underlying distribution is valuable in terms of specific applications, even if it seems not to be a nonparametric approach. In other words the choice of the smoothing parameter in some application can be made by making use of the information known or at least assumed about the distribution form.

Different approaches have been proposed to find a reasonable choice of the h parameter. Among those methods are the least squares cross validation, the likelihood cross validation and the test graph method.

A simulation of a comparative study of some of the kernel methods for sample sizes 25, 50, and 100 for different distributions of varying tail length is presented by Bowman in his 1980 paper.

Mean and Variance Of The Estimator

Theorem

Let K(x) be a symmetric kernel with mean $E_k(x)$ and variance $V_k(x)$ such that $E_k(x)$, $V_k(x) < \infty$ and $\int K(x)dx = 1$. If $\hat{f}(x)$ is a nonparametric kernel estimator based on a sample of size n $(X_1, \cdots, X_n)$ with K(x) as a kernel, then

$\hat{f}(x)$ has a mean $\bar{x}$ and a variance $V_k(x) + s^2$, where $\bar{x}$ is the sample mean and s^2 is the sample variance.

proof

The kernel estimator based on the sample $(X_1, \cdots, X_n)$ is:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{39}$$

Hence the expected value of the random variable x with $\hat{f}(x)$ as a density function will be:

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} \frac{1}{n}\sum_{i=1}^{n} x\frac{K\left(\frac{x-X_i}{h}\right)}{h}dx \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} x\frac{K\left(\frac{x-X_i}{h}\right)}{h}dx \end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$= \bar{x} \tag{40}$$

as for the variance we have:

$$\begin{aligned}
V(x) &= E(x-\bar{x})^2 \\
&= E(x^2) - \bar{x}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} x^2 \frac{K\left(\frac{x-X_i}{h}\right)}{h}dx - \bar{x}^2 \\
&= V_k(x) + \sum_{i=1}^{n}\frac{X_i^2}{n} - \bar{x}^2 \\
&= V_k(x) + s^2
\end{aligned} \tag{41}$$

Corollaries

1. The mean and the variance for a kernel estimator with a Gaussian kernel asymptotically approaches the mean and variance of the empirical distribution function. This can be shown in the following way:

Since in the case of a Gaussian kernel with $\hat{f}(x)$ given as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \tag{42}$$

the expected value and the variance will be:

$$\begin{aligned} E(x) &= \bar{x} \\ V(x) &= h^2 + s^2 \end{aligned} \tag{43}$$

and since h$\longrightarrow$ 0 as n $\longrightarrow \infty$ then E(x)=$\bar{x}$ and V(x) $\longrightarrow s^2$ which are the mean and the variance for the empirical distribution function.

2. For different kernels $K_i(x)$ with variances V_i, i=1,...,n; each used at one of the sample points $X_1, \cdots, X_n$ respectively the mean remains the sample mean while the variance will be:

$$V(x) = \sum_{i=1}^{n} V_i(x) + s^2 \tag{44}$$

The kernel estimator is location and scale invariant and this property is derived in the following section:

Invariance Property Of The Kernel Method

The invariance property for the kernel method is shown in this section under two transformations. First, the location transformation where all the observations are moved either to the left or to the right. Second, the scale transformation where all the observations are either compressed or expanded by a constant factor.

a)The transformation

$$\mathcal{Z}_{\mathrm{i}} = X_{\mathrm{i}} - C \tag{45}$$

$$\hat{f}(x) = \frac{1}{n\,h}\sum_{j=1}^{n}\mathrm{K}\left(\frac{\mathrm{x}-\mathrm{X_j}}{\mathrm{h}}\right) \tag{46}$$

thus

$$\begin{aligned}
\hat{g}(x) &= \frac{1}{n\,h}\sum_{j=1}^{n}\mathrm{K}\left(\frac{\mathrm{x}-\mathrm{Z_j}}{\mathrm{h}}\right) \\
&= \frac{1}{n\,h}\sum_{j=1}^{n}\mathrm{K}\left(\frac{\mathrm{x}+\mathrm{C}-\mathrm{X_j}}{\mathrm{h}}\right) \\
&= \hat{f}(x+C) \qquad (47)
\end{aligned}$$

b)The transformation

$$\mathcal{Z}_{\mathrm{i}} = X_{\mathrm{i}}/\mathrm{k} \tag{48}$$

since the h value is a linear function of the sample standard deviation, hence the new value h value resulting from the transformation of the data by a scale k in the

above way will be $\tilde{h}$ such that:

$$\tilde{h} = \frac{h}{k} \tag{49}$$

thus

$$\begin{aligned}
\hat{g}(x) &= \frac{1}{n\,\tilde{h}} \sum_{j=1}^{n} \mathrm{K}\left(\frac{\mathrm{x} - \mathrm{Z_j}}{\tilde{\mathrm{h}}}\right) \\
&= \frac{1}{n\,\tilde{h}} \sum_{j=1}^{n} \mathrm{K}\left(\frac{\mathrm{x} - \mathrm{X_j/k}}{\tilde{\mathrm{h}}}\right) \\
&= \frac{k}{n\,h} \sum_{j=1}^{n} \mathrm{K}\left(\frac{\mathrm{kx} - \mathrm{X_j}}{\mathrm{h}}\right) \\
&= k\frac{1}{n\,h} \sum_{j=1}^{n} \mathrm{K}\left(\frac{\mathrm{kx} - \mathrm{X_j}}{\mathrm{h}}\right) \\
&= k\,\hat{f}(kx) \qquad (50)
\end{aligned}$$

The Monte Carlo Comparison

The univariate kernel estimate of a density f has the form:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \tag{51}$$

where $X_1, X_2, \ldots, X_n$ are independent, identically distributed observations from f, and K(.) is a function that satisfies the regularity conditions stated on page 19.

The choice of the parameter h in the kernel method is critical, since this parameter controls the smoothness of the resulting estimator.

The choice of the h parameter for the univariate case can frequently be chosen visually in a satisfactory manner (Wahba, 1983). However, the need for a predetermined choice of the h parameter in most of the applications of the nonparametric density estimation suggests the idea of examining the behavior of some of the different distributions under a proposed choice of h. A Monte Carlo experiment of size 1000 is used to examine the behavior of the estimators for six different distributions. These distributions are:

-Uniform.

-Exponential.

-Cauchy.

-Double Exponential.

-Logistic.

-Normal.

The criteria chosen for the comparison is the mean integrated square error.

The optimum choice for h is shown in equation (38) to be a constant times $n^{-\frac{1}{5}}$. Furthermore, Silverman (1986) shows that the optimum h for the normal is

$1.06\sigma n^{-\frac{1}{5}}$ using a normal kernel. Therefore, a data dependent h equals to $sn^{-\frac{1}{5}}$, where s represents the sample standard deviation for sample size n, is chosen for the Monte Carlo experiment study. It also gives a scale invariant nonparametric density estimate since s is a scale incariant estimator of σ.

Now, the data based choice of the h was used with the kernel technique when a Gaussian kernel is utilized in which case the estimator will take the form:

$$\hat{f}(x) = \frac{1}{n\,h}\sum_{j=1}^{n}\phi\left(\frac{x - X_j}{h}\right) \tag{52}$$

where $\phi(x)$ represents the p.d.f for the standard normal distribution. Sample sizes 10,20,...,60 i.e 10(10)60 are used and MISE defined as:

$$\begin{aligned} MISE &= \int E\left[\hat{f}(x) - f(x)\right]^2 dx \\ &= E\int\left[\hat{f}(x) - f(x)\right]^2 dx \end{aligned} \tag{53}$$

where $\hat{f}(x)$ denotes the nonparametric estimator based on the previous choice of h. while $f(x)$ will be one of the mentioned six distributions.

To evaluate the performance of the method over the various distributions, the Monte Carlo experiment is designed the same way for all the six distributions and the different sample sizes.

The methodology is such that a certain sample size 10(10)60 of each of the distributions is generated using the IMSL routines RNUN, RNEXP, RNCAU, RNNOR

for the uniform, exponential, cauchy and normal distributions respectively. While an inverse C.D.F technique is used for the double exponential and the logistic distributions. The data based choice of the smoothing parameter is then calculated for each of the 1000 different samples. The integrated square error ISE given as:

$$ISE = \int \left[\hat{f}(x) - f(x)\right]^2 dx \tag{54}$$

is then computed for each sample using the IMSL integration routine QDAGI with bounds $-\infty$ and ∞. This is only modified to be (-50 , +50) for the logistic distribution to avoid the numerical difficulty of computation beyond this limits. An estimate of MISE is then obtained by averaging the ISE from the 1000 Monte Carlo repetitions. Likewise, an estimate of the standard deviation of MISE is computed. The results of the Monte Carlo experiment for the different sample sizes are given in Table2 where the table entries give the MISE for different sample sizes with the standard deviation in brackets.

The results of the Monte Carlo show that the choice of h which is near optimal for the normal (h_{opt} for the normal is $1.06\sigma n^{-\frac{1}{5}}$) gives a comparable results for the double exponential and the logistic distributions, while a reasonable fit was found for the Cauchy. A relatively large MISE is obtained for the uniform and exponential distributions which indicates that the choice for these distributions is not as optimal.

Table 2. Values of MISE for different distributions with standard deviation based on M.C size 1000 and sample of size n for each repetition

n	*Uniform*	*Expon.*	*Cauchy*	*D.E*	*Logistic*	*Normal*
10	0.19142 *(0.13260)*	0.15297 *(0.05475)*	0.06277 *(0.03842)*	0.04318 *(0.02920)*	0.06412 *(0.06444)*	0.03004 *(0.02561)*
20	0.12664 *(0.05745)*	0.12950 *(0.03629)*	0.06834 *(0.04147)*	0.02672 *(0.01526)*	0.04487 *(0.03402)*	0.01970 *(0.01617)*
30	0.10690 *(0.03691)*	0.11982 *(0.02800)*	0.06991 *(0.04038)*	0.02140 *(0.01146)*	0.04131 *(0.02593)*	0.01407 *(0.01042)*
40	0.09509 *(0.03028)*	0.11237 *(0.02478)*	0.07368 *(0.04011)*	0.01835 *(0.00961)*	0.03989 *(0.02234)*	0.01154 *(0.00843)*
50	0.08800 *(0.02429)*	0.10575 *(0.02220)*	0.07727 *(0.03953)*	0.01651 *(0.00825)*	0.03916 *(0.01934)*	0.00996 *(0.00704)*
60	0.08267 *(0.02017)*	0.10191 *(0.01889)*	0.07998 *(0.03921)*	0.01503 *(0.00731)*	0.03904 *(0.01763)*	0.00883 *(0.00625)*

IV. Optimal Choice Of The Smoothing Parameter

Introduction

The choice of the h parameter is the most essential step in successful non-parametric density estimation using the kernel method. This choice is theoretically derived based on the optimization of the approximated MISE defined in chapter III. In this chapter a Monte Carlo experiment is performed to approximate the optimal choice of the h parameter for the Gaussian kernel. This choice is a crucial one in terms of a goodness of fit application besides any other applications that require a nonparametric estimate of the density. The different distributions considered are the:

- Uniform
- Exponential
- Cauchy
- Double Exponential
- Logistic
- Normal

These distributions represent different shapes and characteristics.

Hence, the purpose of this chapter is to find an optimal h and the corresponding MISE for a 1000 different samples each of size 20 from the above distributions.

Methodology

The optimal h w.r.t minimizing the approximate MISE is given as:

$$h_{opt} = m_2^{-2/5} \left\{ \int K^2(t)dt \right\}^{1/5} \left\{ \int f''(x)^2 \, dx \right\}^{-1/5} n^{-1/5} \tag{55}$$

where m_2 is the kernel second moment (see Parzen, 1962).

This approximate optimal value h_{opt} is derived for the different distributions as a first step:

1. Uniform distribution

For the uniform distribution the approximate expression gives a zero h since the density is constant. This case corresponds to the E.D.F estimator. However the M.C results indicate that this value does differ from zero.

2. Exponential distribution

For the one parameter exponential distribution with variance $V(x) = \frac{1}{\alpha^2}$ and p.d.f given as:

$$f(x) = \alpha \, e^{-\alpha x} \quad , \alpha > 0 \; , x \geq 0 \tag{56}$$

$$f'(x) = -\alpha^2 \, e^{-\alpha x} \tag{57}$$

$$f''(x) = \alpha^3 \, e^{-\alpha x} \tag{58}$$

$m_2 = 1$ for the Gaussian kernel

$$\begin{aligned}
\int K^2(t)dt &= \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi}} exp\left(\frac{-x^2}{2}\right)\right]^2 \\
&= \frac{1}{2\sqrt{\pi}} \\
&= .2821 \qquad (59) \\
\int_0^{\infty} f''^2(x)dx &= \alpha^6 \int_0^{\infty} e^{-2\alpha x} dx \\
&= \frac{\alpha 5}{2} \\
&= \frac{1}{2\,\sigma^5} \qquad (60)
\end{aligned}$$

where σ is the standard deviation of the distribution. Hence by substituting in the formula for the approximate optimal h, the corresponding h for this distribution will be:

$$h_{opt} = .8918\sigma\, n^{-\frac{1}{5}} \qquad (61)$$

3. Cauchy distribution

For the Cauchy distribution with a density f(x), the optimal h is derived below:

$$\begin{aligned}
f(x) &= \frac{1}{\pi\left[(x-a)^2+1\right]} \quad -\infty < x < \infty, -\infty < a < \infty \qquad (62) \\
f''(x) &= \frac{2\left[3\pi\,(x-a)^2-1\right]}{\pi^2\left[(x-a)^2+1\right]^3} \qquad (63)
\end{aligned}$$

$$
\begin{aligned}
\int_{-\infty}^{\infty} f''^2(x)dx &= \int_{-\infty}^{\infty} \left[\frac{2\left[3\pi\,(x-a)^2 - 1\right]}{\pi^2\left[(x-a)^2+1\right]^3} \right]^2 dx \\
&= \int_{-\infty}^{\infty} \left[\frac{2\,[3\pi y^2 - 1]}{\pi^2\,[y^2+1]^3} \right]^2 dy \\
&= .0411 \left[\int_{-\infty}^{\infty} \frac{9\pi^2 y^4\, dy}{[y^2+1]^6} - \int_{-\infty}^{\infty} \frac{6\pi y^2\, dy}{[y^2+1]^6} + \int_{-\infty}^{\infty} \frac{dy}{[y^2+1]^6} \right] \\
&= (.0411)(2) \left[\int_{0}^{\infty} \frac{9\pi^2 y^4\, dy}{[y^2+1]^6} - \int_{0}^{\infty} \frac{6\pi y^2\, dy}{[y^2+1]^6} + \int_{0}^{\infty} \frac{dy}{[y^2+1]^6} \right] \\
&= .0411 \left[18\pi^2 B\left(\frac{5}{2}, \frac{7}{2}\right) - 12\pi B\left(\frac{3}{2}, \frac{9}{2}\right) + 2B\left(\frac{1}{2}, \frac{11}{2}\right) \right] \\
&= .0411 \left[177.699 B\left(\frac{5}{2}, \frac{7}{2}\right) - 37.699 B\left(\frac{3}{2}, \frac{9}{2}\right) + 2B\left(\frac{1}{2}, \frac{11}{2}\right) \right] \\
&= .1992 \qquad (64) \\
h_{opt} &= 1.0721\, n^{-\frac{1}{5}} \qquad (65)
\end{aligned}
$$

where B is a beta function.

4. Double Exponential distribution

For the double exponential density given by:

$$f(x) = \frac{e^{-\frac{|x-\theta|}{\beta}}}{2\beta} \quad , \beta > 0 \tag{66}$$

and similar to the previous case the optimal approximate h is:

$$h_{opt} = .7244\sigma n^{-\frac{1}{5}} \tag{67}$$

5. Logistic distribution

The logistic density function in two parameters is given by:

$$f(x) \; = \; exp\left[-\left(x-a\right)/b\right] / \left[b\left(1+exp\left[-\left(x-a\right)/b\right]\right)^2\right] \tag{68}$$

with

E(x) = a , V(x) = $\frac{(b\pi)^2}{3}$, mode(x) = a

$$\begin{aligned}
f'(x) &= \frac{\frac{1}{b}e^{\frac{x-a}{b}}\left[b\left(1+e^{\frac{x-a}{b}}\right)^2\right] - 2b\left(1+e^{\frac{x-a}{b}}\right).\frac{1}{b}e^{2\left(\frac{x-a}{b}\right)}}{b^2\left(1+e^{\frac{x-a}{b}}\right)^4} \\
&= \frac{e^{\frac{x-a}{b}}\left(1+e^{\frac{x-a}{b}}\right)^2 - 2\left(1+e^{\frac{x-a}{b}}\right).e^{2\left(\frac{x-a}{b}\right)}}{b^2\left(1+e^{\frac{x-a}{b}}\right)^4} \\
&= \frac{e^{\frac{x-a}{b}}\left(1+e^{\frac{x-a}{b}}\right) - 2.e^{2\left(\frac{x-a}{b}\right)}}{b^2\left(1+e^{\frac{x-a}{b}}\right)^3} \\
&= \frac{e^{\frac{x-a}{b}}\left(1-e^{\frac{x-a}{b}}\right)}{b^2\left(1+e^{\frac{x-a}{b}}\right)^3}
\end{aligned} \tag{69}$$

$$
\begin{aligned}
f''(x) &= \frac{\frac{1}{b}e^{\frac{x-a}{b}}\left(1-2e^{\frac{x-a}{b}}\right)b^2\left(1+e^{\frac{x-a}{b}}\right)^3 - 3b^2\left(1+e^{\frac{x-a}{b}}\right)^2\frac{1}{b}e^{\frac{x-a}{b}}e^{\frac{x-a}{b}}\left(1-e^{\frac{x-a}{b}}\right)}{b^4\left(1+e^{\frac{x-a}{b}}\right)^6} \\
&= \frac{e^{\frac{x-a}{b}}\left(1-2e^{\frac{x-a}{b}}\right)\left(1+e^{\frac{x-a}{b}}\right) - 3e^{2\left(\frac{x-a}{b}\right)}\left(1-e^{\frac{x-a}{b}}\right)}{b^3\left(1+e^{\frac{x-a}{b}}\right)^4} \\
&= \frac{e^{\frac{x-a}{b}}\left[1-e^{\frac{x-a}{b}}-2e^{2\left(\frac{x-a}{b}\right)}-3e^{\frac{x-a}{b}}+3e^{2\left(\frac{x-a}{b}\right)}\right]}{b^3\left(1+e^{\frac{x-a}{b}}\right)^4} \\
&= \frac{e^{\frac{x-a}{b}}\left[1-4e^{\frac{x-a}{b}}+e^{2\left(\frac{x-a}{b}\right)}\right]}{b^3\left(1+e^{\frac{x-a}{b}}\right)^4} \qquad (70)
\end{aligned}
$$

Now, let $y = exp(\frac{x-a}{b})$ and hence

$$
\begin{aligned}
\int_{-\infty}^{\infty} f''(x)dx &= \int_0^{\infty} \frac{y^2(1-4y+y^2)^2}{b^6(1+y)^8}\frac{b}{y}dy \\
&= \frac{1}{b^5}\int_0^{\infty}\frac{y-8y^2+18y^3-8y^4+y^5}{(1+y)^8}dy \\
&= \frac{1}{b^5}\left[B(2,6)-8B(3,5)+18B(4,4)-8B(5,3)+B(6,2)\right] \\
&= \frac{1}{b^5}\left[2B(2,6)-16B(3,5)+18B(4,4)\right] \\
&= \frac{1}{b^5}(.02381) \qquad (71)
\end{aligned}
$$

Thus, the optimal h for the Gaussian kernel is given as:

$$
\begin{aligned}
h_{opt} &= 1.6396\, b\, n^{-\frac{1}{5}} \\
&= \frac{\sqrt{3}}{\pi}(1.6396)\,\sigma\, n^{-\frac{1}{5}} \\
&= 0.9039\,\sigma\, n^{-\frac{1}{5}} \qquad (72)
\end{aligned}
$$

whereσ is the distribution standard deviation.

6. Normal distribution

Silverman (1986) shows, as an example, that the value of h based on the previous approximation is:

$$h_{opt} = 1.06\sigma n^{-\frac{1}{5}} \tag{73}$$

The h value obtained is summarized in the following table (Table3.).

Table 3. h Values for Different Distributions

Distribution	h
Exponential	$.8918\ \sigma n^{-\frac{1}{5}}$
Cauchy	$1.0721\ n^{-\frac{1}{5}}$
Double Exponential	$.7244\ \sigma n^{-\frac{1}{5}}$
Logistic	$.9039\ \sigma n^{-\frac{1}{5}}$
Uniform	0.0
Normal	$1.06\ \sigma n^{-\frac{1}{5}}$

A numerical improvement of the previous recommended h is to be found based on the use of an unbiased estimator for the standard deviation and the use of a linear search around the previous h value for an h with smaller MISE. The general form of the proposed estimator for h is a constant multiple of the unbiased estimator of σ times $n^{-\frac{1}{5}}$. Let $\tilde{\sigma}$ represents the unbiased estimator of σ, which is given as:

$$\tilde{\sigma} = \hat{\sigma}\frac{\sqrt{n}\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)} \tag{74}$$

where Γ is the gamma function and $\hat{\sigma}$ is given by:

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \tag{75}$$

thus

$$\begin{aligned} \tilde{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\,\frac{\sqrt{n}\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)} \\ &= \frac{\sqrt{n-1}\,S\,\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)} \end{aligned} \tag{76}$$

The Monte Carlo experiment here is designed for a sample size 20 in which case the optimal h is assumed to be $h_{opt} = k\tilde{\sigma}n^{-\frac{1}{5}}$. The experiment starts with generating 1000 samples from the 6 distributions (uniform, exponential, Cauchy, double exponential, logistic, and normal) each of size 20. Defining an interval $\mathcal{H}$ such that:

$$\mathcal{H} = \{h1 | h_{opt} - l \leq h1 \leq h_{opt} + u\} \tag{77}$$

where h_{opt} is as defined in the above table, a search in the closed interval $\mathcal{H}$ for an h1 that minimizes the MISE is performed. The search starts by subdividing the interval into a mesh of m equal subintervals. Computing the MISE at each of (m+1) end points of the subintervals gives an array of MISE's. The minimum MISE corresponds to an optimal h1 value in $\mathcal{H}$. If the optimal h1 lies on either end of the interval $\mathcal{H}$ then the search interval is expanded by l or u for lower or upper end points of $\mathcal{H}$

respectively and the search continues.

Upon finding the optimal h1, as described above, a constant k is computed as:

$$k = \frac{h1}{\tilde{\sigma} n^{-\frac{1}{5}}} \tag{78}$$

which defines the factor that relates the choice of the h1 to the unbiased estimator of the standard deviation and the sample size. The following table gives the average optimal h1, the average k, and the average MISE over the 1000 different samples with their standard deviations in brackets.

Table 4. Optimal h for sample sizes 20 for Different Distributions

Distribution	# of samples	$h1_{opt}$	k	MISE
Uniform	698	.1629	1.0589	.1116
		(.0515)	*(.4443)*	*(.0394)*
Exponential	163	.2719	.5334	.0865
		(.0915)	*(.1902)*	*(.0304)*
Cauchy	163	7.1591	.9657	.0675
		(12.8467)	*(.0800)*	*(.0407)*
Double Exponential	600	.5922	.8376	.0216
		(.1141)	*(.2723)*	*(.0140)*
Logistic	347	.7862	1.4821	.0211
		(.0899)	*(.1142)*	*(.0164)*
Normal	1000	.6160	1.1789	.0145
		(.1008)	*(.3190)*	*(.0124)*

The method shows an improvement over the choice of h as $sn^{-\frac{1}{5}}$. The percentage improvement for each distribution is given in the following table:

Table 5. Percentage improvement in MISE relative to choice of h as $sn^{-\frac{1}{5}}$ for different distributions

Distribution	% improvement
Uniform	11
Exponential	33.2
Cauchy	1
Double Exponential	19
Logistic	52
Normal	26.4

This shows that the choice of $sn^{-\frac{1}{5}}$ is rather good one over the set of distributions studied.

The following graphs show examples from uniform, normal, exponential, and logistic distributions using the constant k given in table 4.

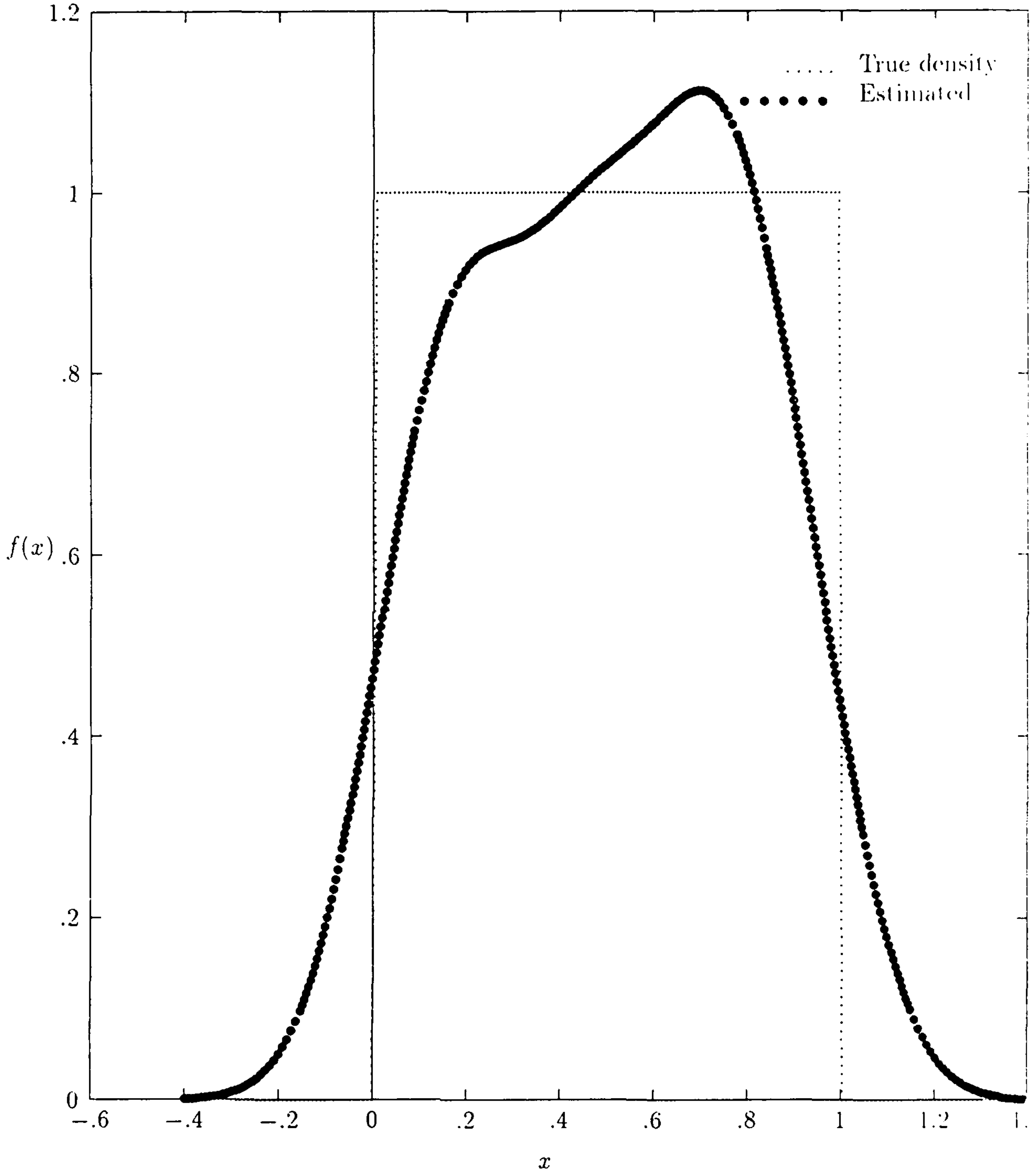

Figure 1. A nonparametric p.d.f. for the uniform distribution with sample size 60

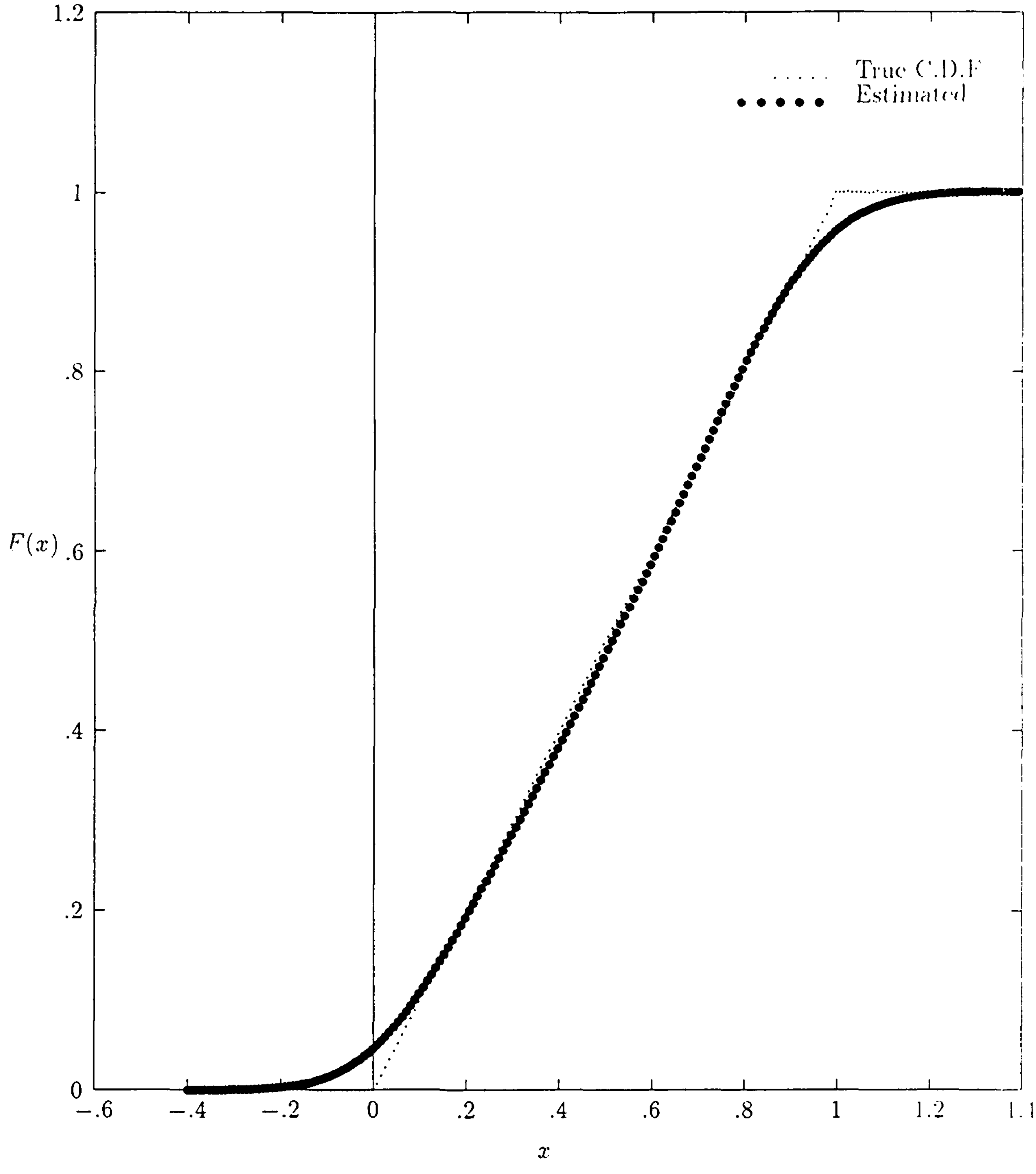

Figure 2. A nonparametric c.d.f. for the uniform distribution with sample size 60

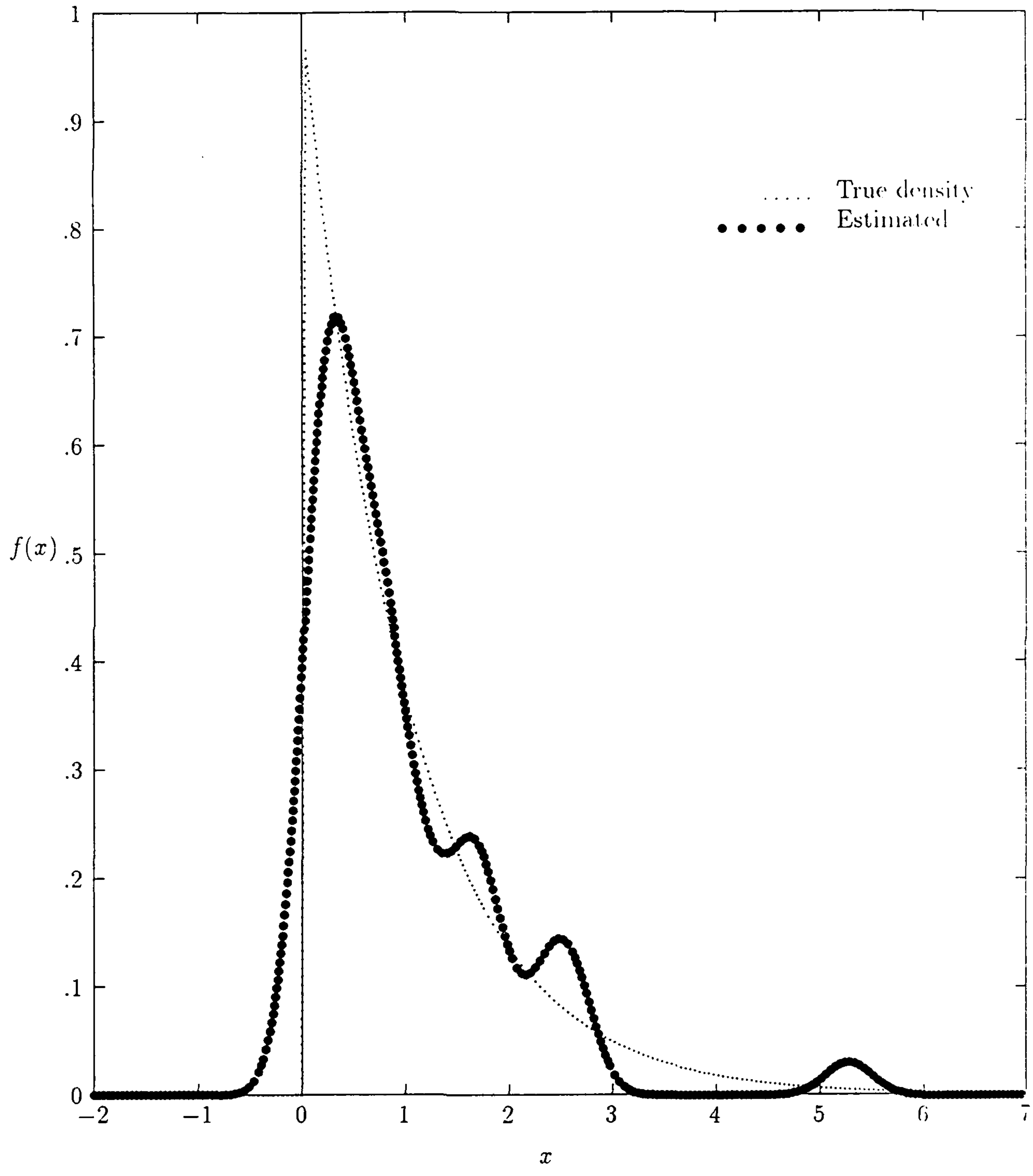

Figure 3. A nonparametric p.d.f. for the exponential distribution with sample size 60

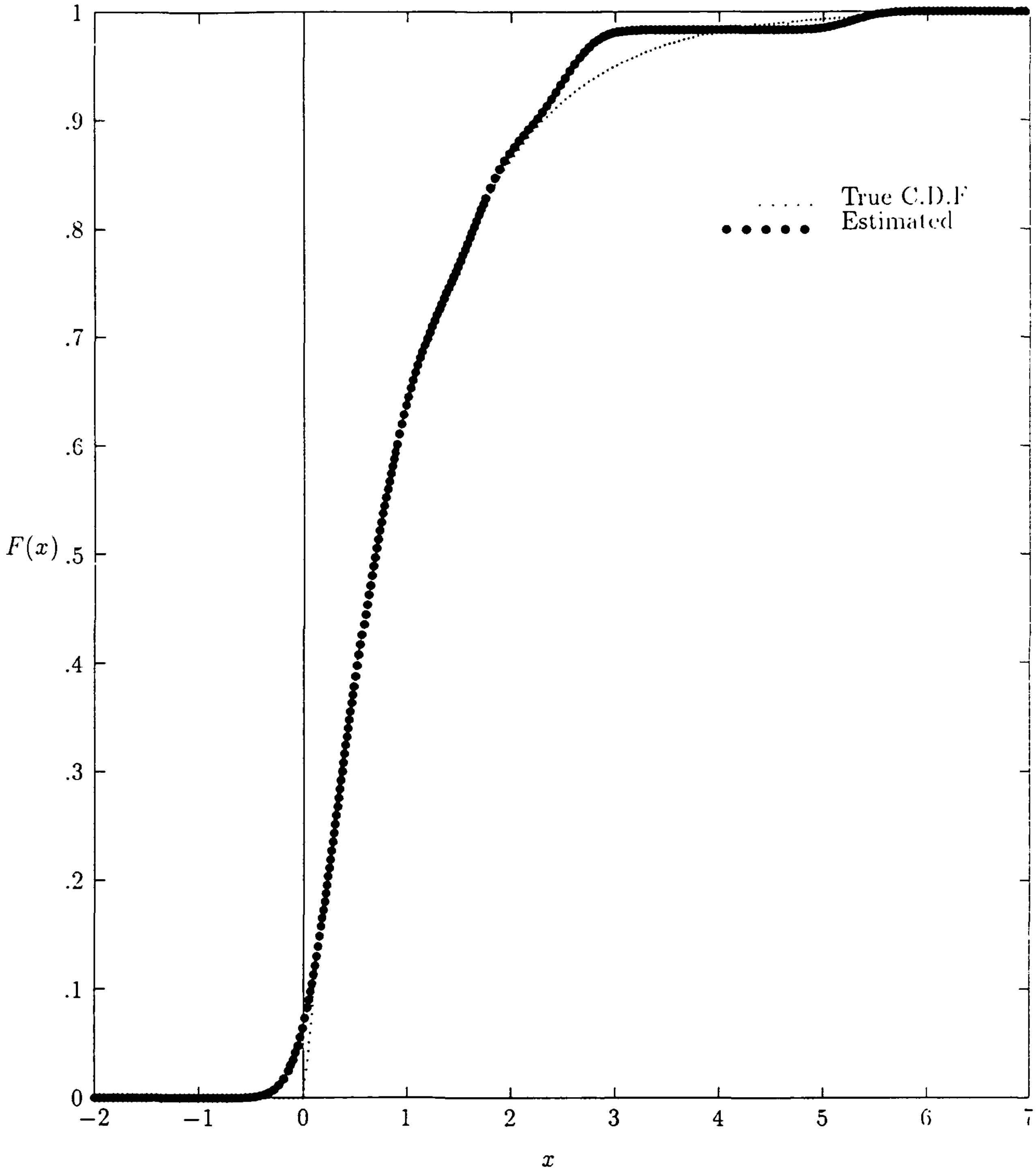

Figure 4. A nonparametric c.d.f. for the exponential distribution with sample size 60

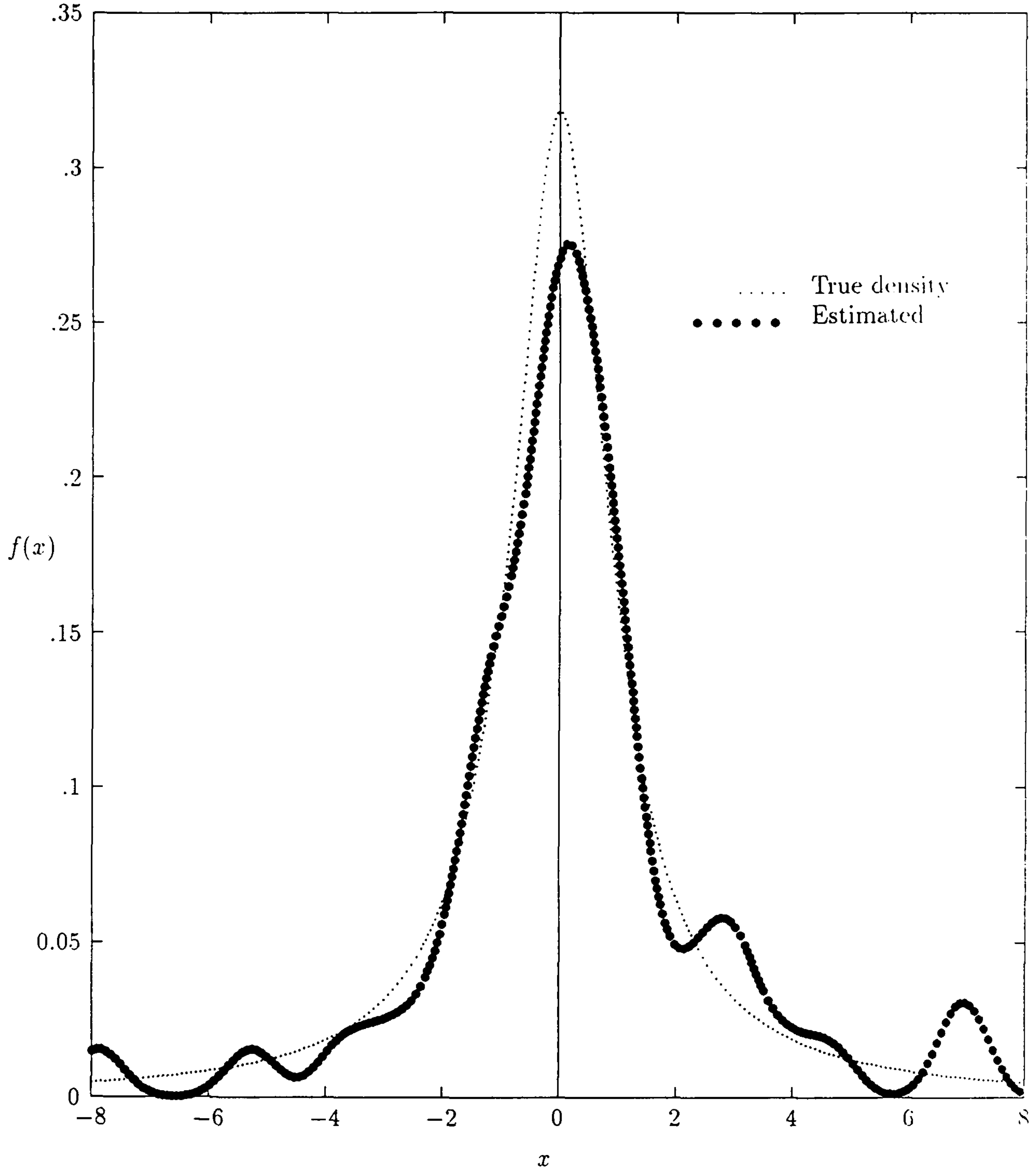

Figure 5. A nonparametric p.d.f. for the Cauchy distribution with sample size 60

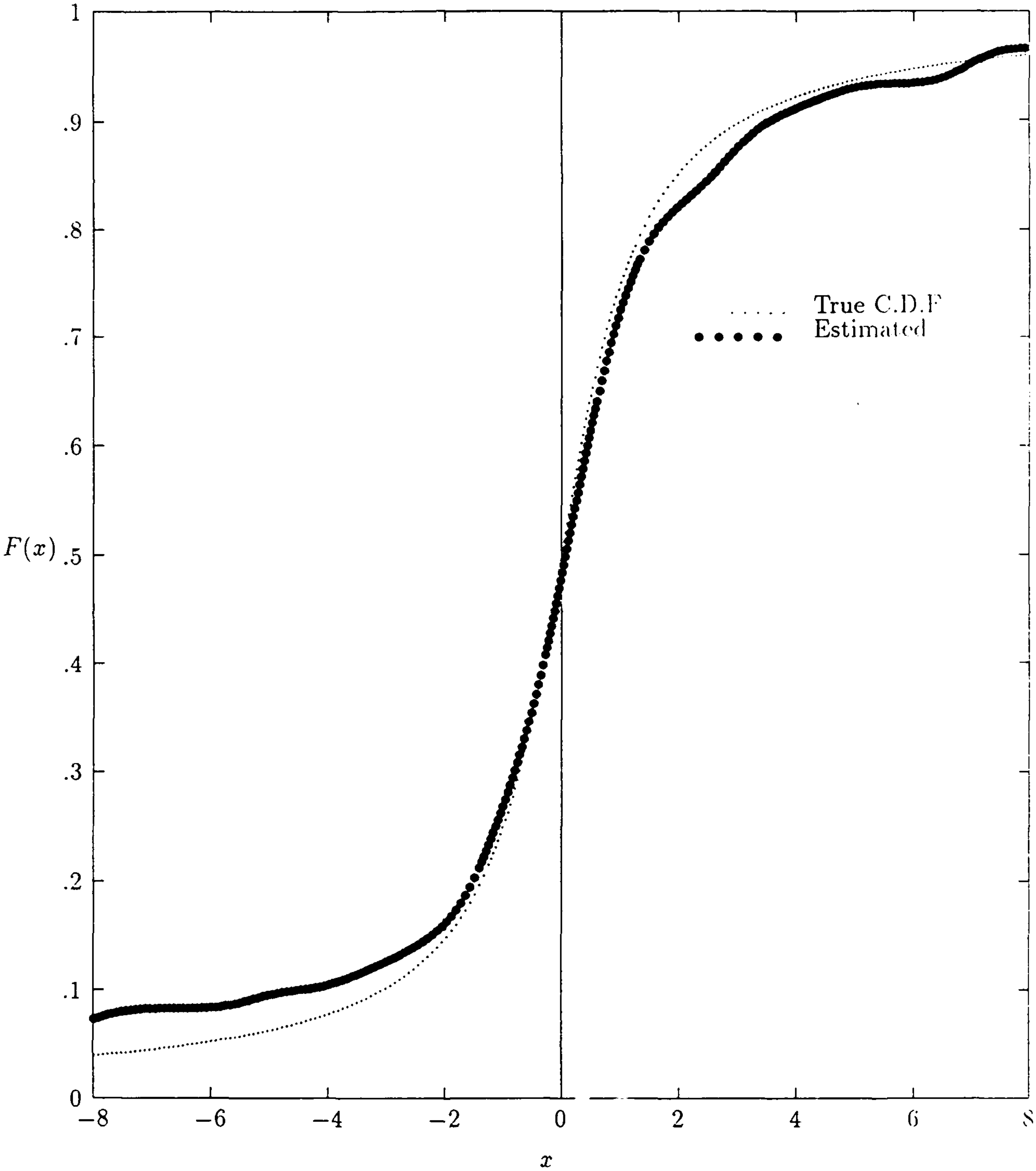

Figure 6. A nonparametric c.d.f. for the Cauchy distribution with sample size 60

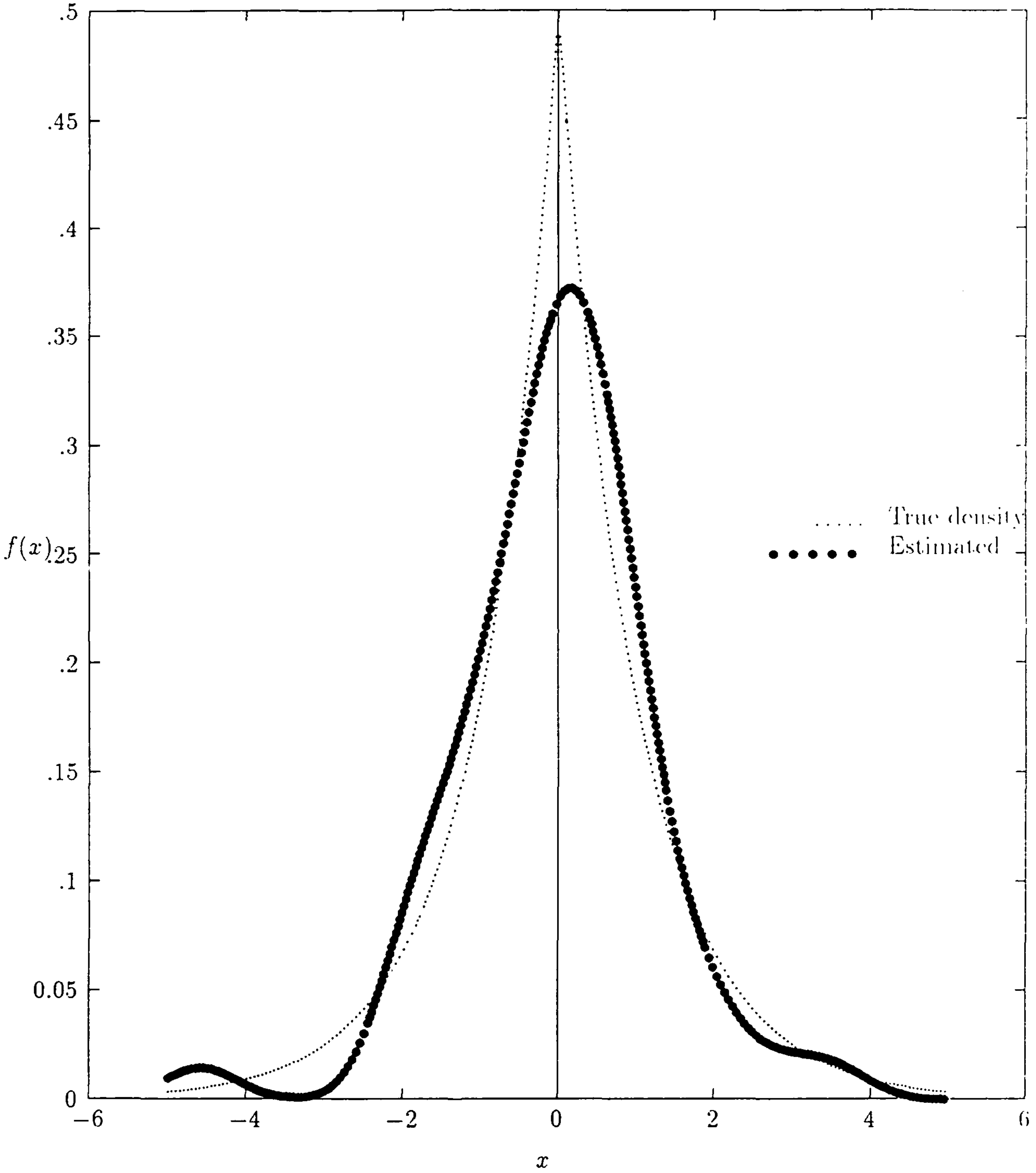

Figure 7. A nonparametric p.d.f. for the double exponential distribution with sample size 60

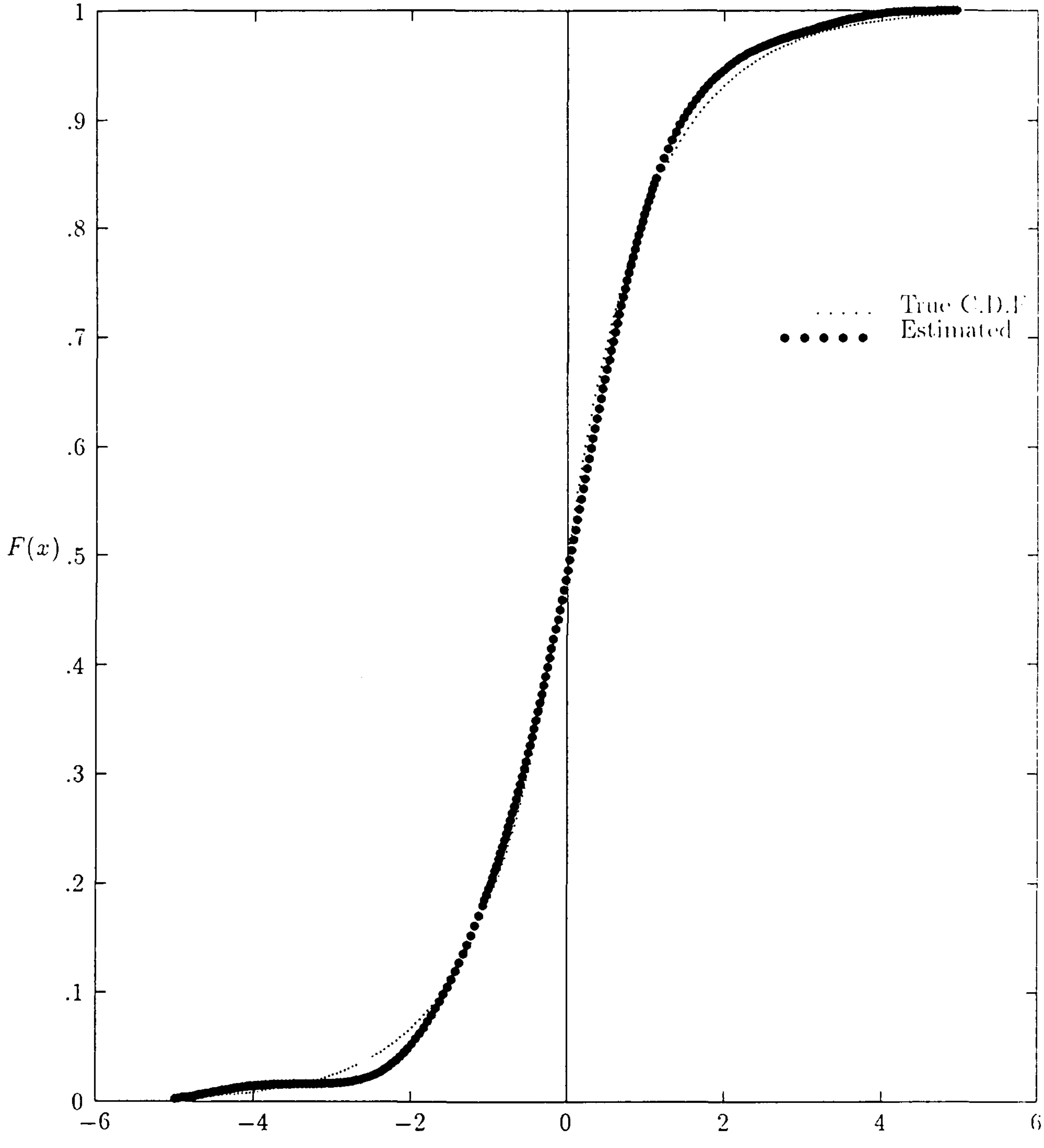

Figure 8. A nonparametric c.d.f. for the double exponential distribution with sample size 60

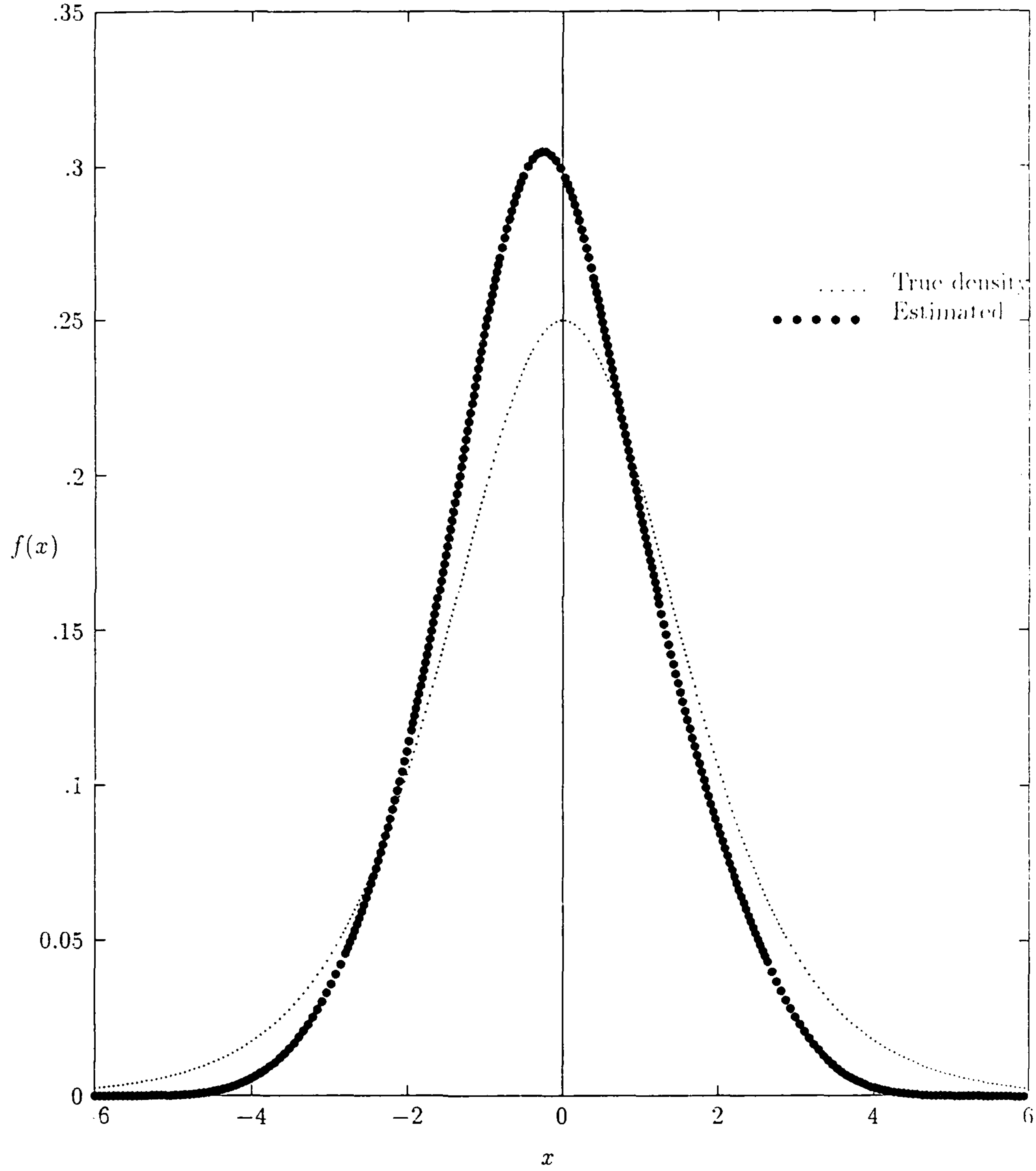

Figure 9. A nonparametric p.d.f. for the logistic distribution with sample size 60

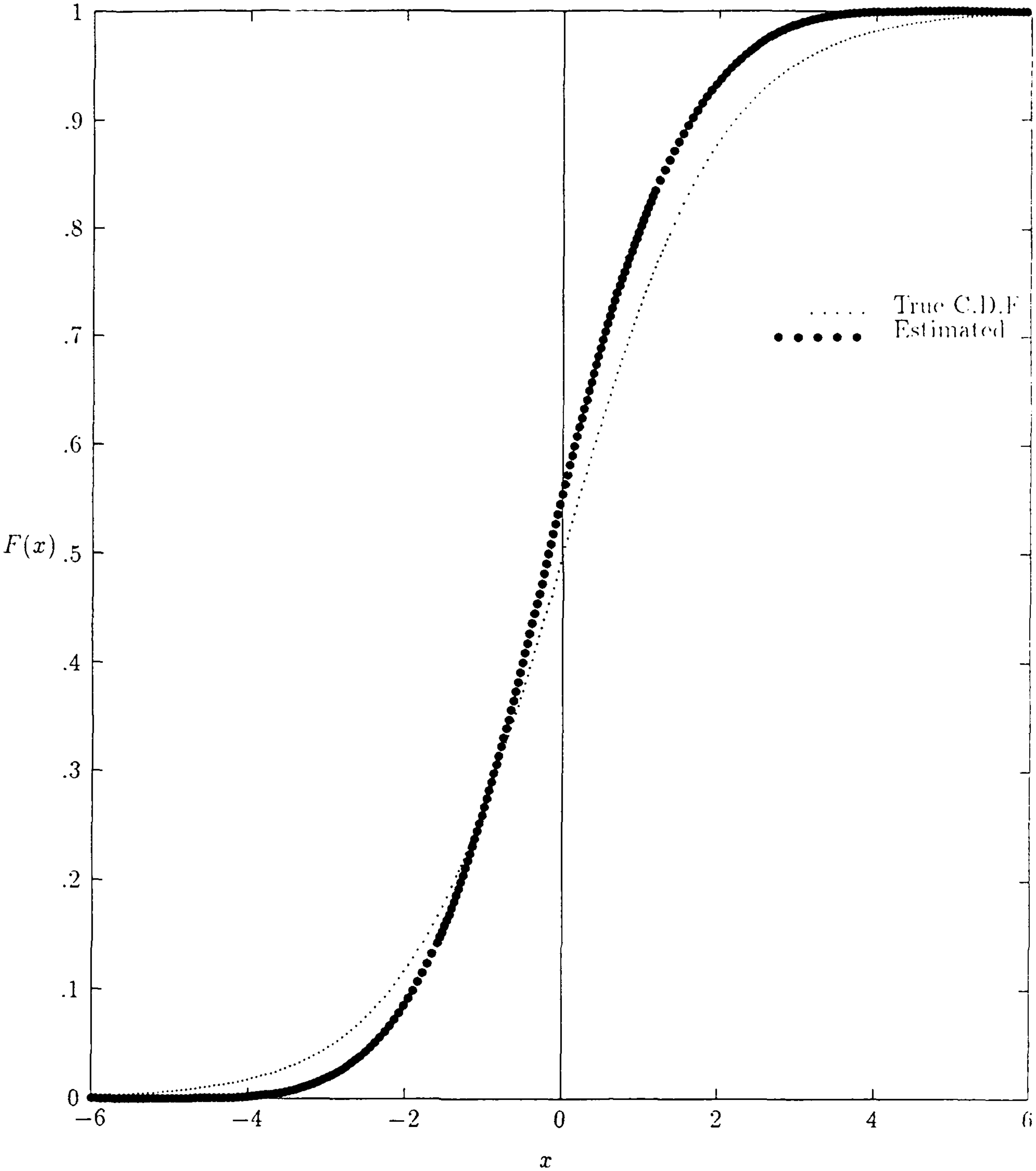

Figure 10. A nonparametric c.d.f. for the logistic distribution with sample size 60

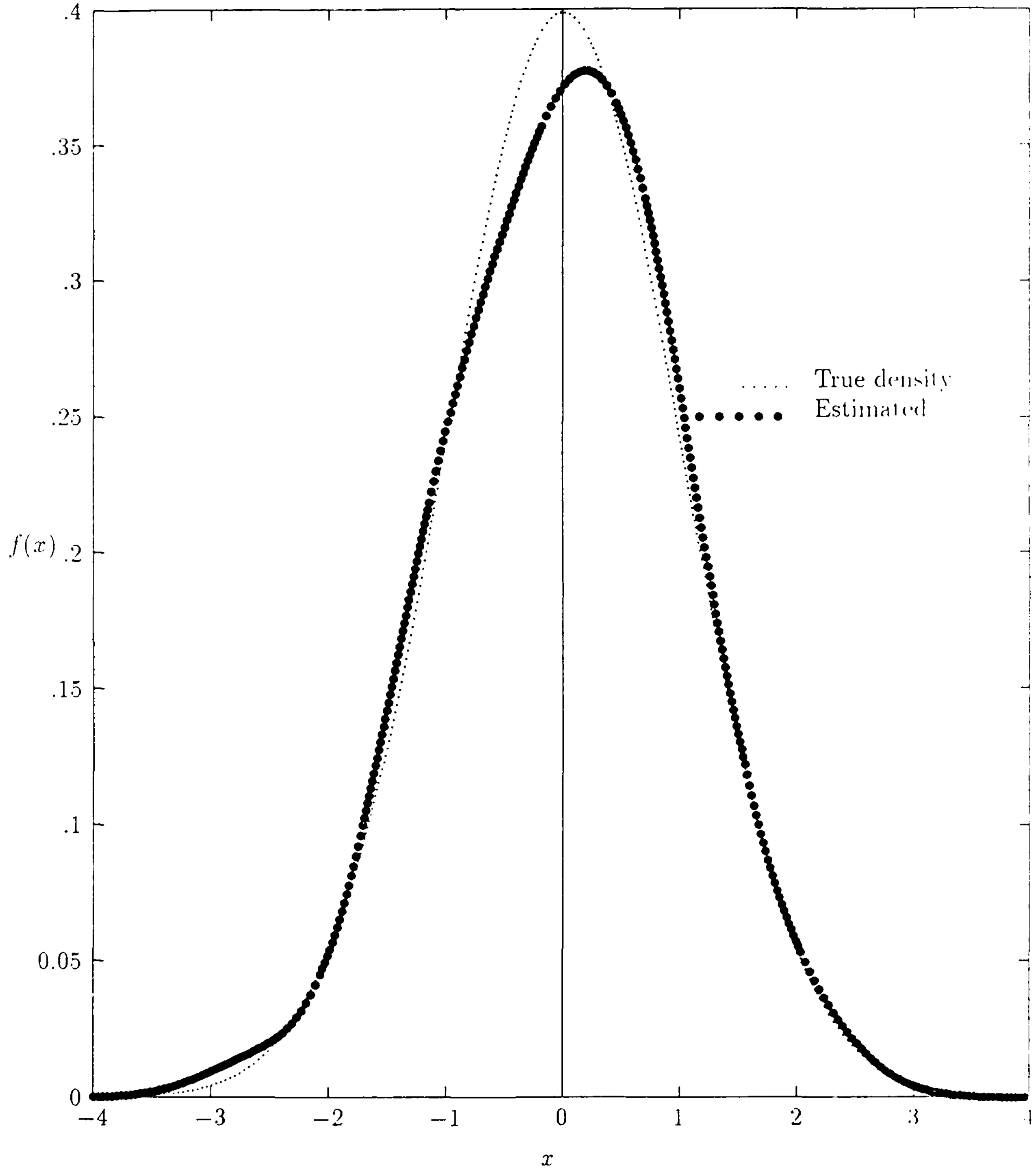

Figure 11. A nonparametric p.d.f. for the normal distribution with sample size 60

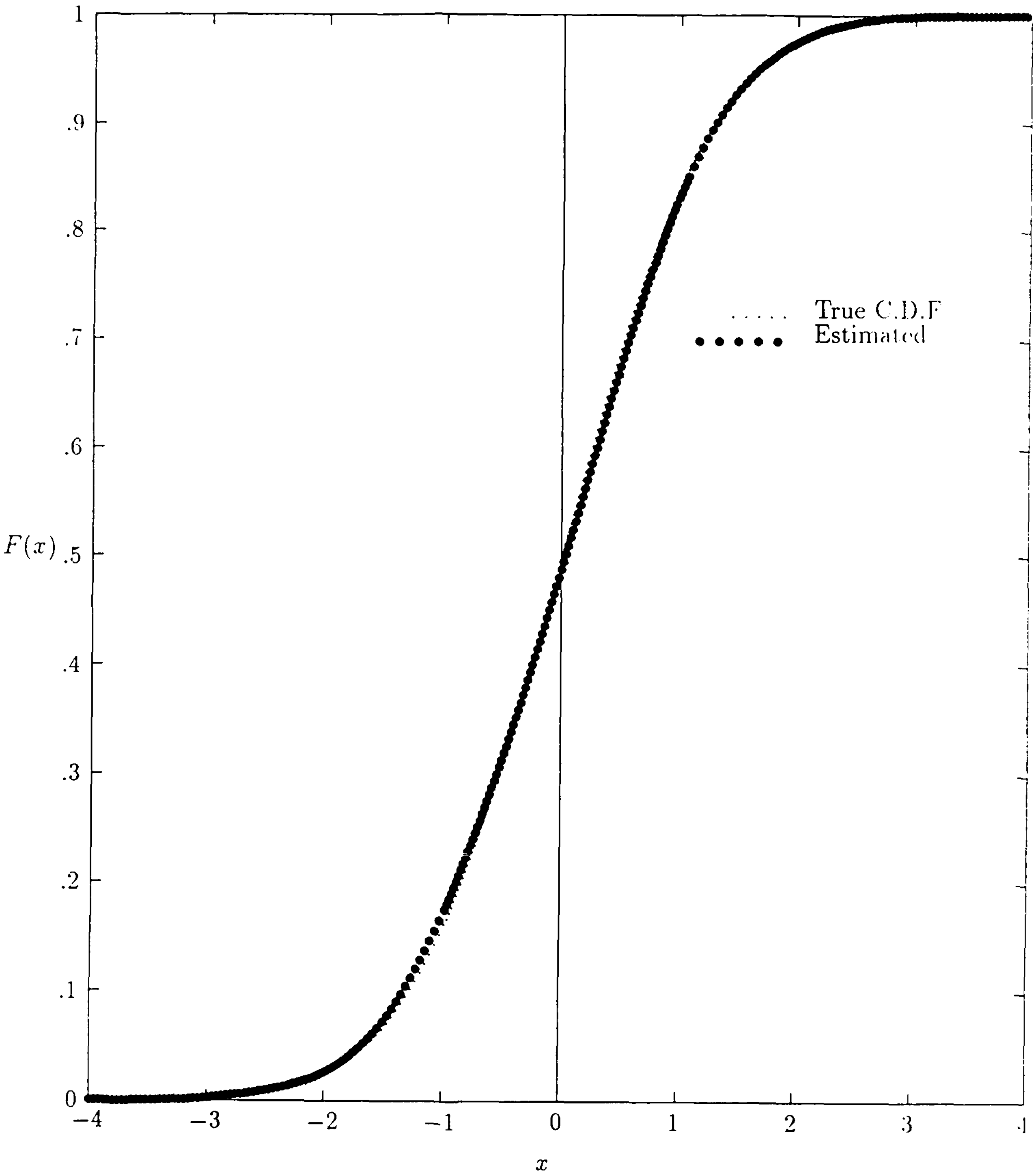

Figure 12. A nonparametric c.d.f. for the normal distribution with sample size 60

Next, an example from each distribution is given for a sample of size 60. The h parameter used is $h = ksn^{-\frac{1}{5}}$. The seed used for the uniform distribution is the same for the other distributions. The value of the ISE for the uniform distribution is .0514, for the Cauchy is .0632, for the double exponential is .0076, for the logistic is .0064, and for the normal is .0012.

The uniform distribution fit shows an almost linear behavior for the C.D.F. in the interval [.1,.9], however due to the infinite support of the Gaussian kernel. the support of the estimated density is [-.4,1.4]. This is a typical behavior for such estimator and remedial measures can be taken to handle such a case, however the objective is to get a tool that can be used in Monte Carlo of relatively large size where it is not possible to visually examine each case separately. The exponential distribution example shows that the estimated density support is close to the real support. It also indicates that the estimated density is not quite smooth. The behvior can be improved by a larger choice of the h parameter, however this causes the ISE to be larger. The bump near x=5.5 is due to the existence of at least an observation near the upper tail portion of the support. For the Cauchy distribution a constant multiple of $n^{-\frac{1}{5}}$ is used as pointed in the Monte Carlo experiment. Both tails are rough, however the middel portion of the distribution is reasonalbely close. The double exponential case gives a fairly close fit except at the lower tail of the distribution. The logistic distribution, for this case, does not give as good fit as for the normal. The normal distribution example shows a good fit at both tails with the

noparametric distribution skewed to the right. This indicates that the observations from the sample selected are not quite symmetric about the true mean and hence it shows how the nonparametric distribution follows the sample behavior.

V. Parameter Estimation

Introduction

Parameter estimation for the three parameter Weibull distribution is discussed in this chapter. While the problem has been handled in different ways, the method used here is based on the numerical solution of the log-likelihood equations using the hybrid method. The hybrid method is an iterative method. The method is surveyed and the stopping rule is stated. The results from this chapter will be compared with the results from the next chapter.

Maximum Likelihood Estimation For The Parameters Of The Three Parameter Weibull Distribution

The likelihood function for the three parameters weibull is given by:

$$
\begin{aligned}
L(x_1, \cdots, x_n, \delta, \theta, \beta) &= \prod_{i=1}^{n} f(x_i; \delta; \theta; \beta) \\
&= \left(\beta e^{-\beta}\right)^n \prod_{i=1}^{n} \left[(x_i - \delta)^{\beta - 1} \, exp\left(-\theta^{-\beta} \sum_{i=1}^{n} (x_i - \delta)^{\beta} \right) \right] \quad (79)
\end{aligned}
$$

Which gives the following set of equations upon the differentiation of the log likelihood function w.r.t the three unknown parameters.

$$(-n\beta/\theta) + \beta e^{-(\beta+1)} \sum_{i=1}^{n} (x_i - \delta)^{\beta} = 0 \quad (80)$$

$$(n/\beta) - n ln\theta + \sum_{i=1}^{n} ln\,(x_i - \delta) + \theta^{-\beta} ln\theta \sum_{i=1}^{n} (x_i - \delta)^{\beta}$$

$$-\theta^{-\beta} \sum_{i=1}^{n} \left[(x_i - \delta)^{\beta}\, ln\,(x_i - \delta) \right] = 0 \quad (81)$$

$$-(\beta - 1) \sum_{i=1}^{n} (x_i - \delta)^{-1} + \beta e^{-\beta} \sum_{i=1}^{n} (x_i - \delta)^{\beta - 1} = 0 \quad (82)$$

The solution of these equations gives a vector $\hat{\Theta}_l = (\hat{\delta}, \hat{\theta}, \hat{\beta})$ that maximizes the log likelihood function (also, maximizes the likelihood function).

The first equation gives the parameter θ as a function of δ and β in the form:

$$\theta = \theta(\delta, \beta) \quad (83)$$

while the other two equations are not explicitly solvable for β, δ. By substituting θ from the first equation into the other two equations, these last two equations become:

$$\frac{1}{\beta}+\frac{1}{n}\sum_{i=1}^{n} ln\,(x_i-\delta)-\left[\sum_{i=1}^{n}(x_i-\delta)^{\beta}\right]^{-1}\left[\sum_{i=1}^{n}(x_i-\delta)^{\beta}\,ln\,(x_i-\delta)\right] = 0 \quad (84)$$

$$-(1-\beta)\sum_{i=1}^{n}(x_i-\delta)^{-1}+n\beta\left[\sum_{i=1}^{n}(x_i-\delta)^{\beta}\right]^{-1}\left[\sum_{i=1}^{n}(x_i-\delta)^{\beta-1}\right] = 0 \quad (85)$$

The system of the 3 non linear equations for the maximum likelihood in $\Theta = (\delta, \theta, \beta)$ is solved using a numerical technique. The method is known as the hybrid method. This method is basically an iterative method based on Newton-Raphson method, where the equations have the form:

$$L_i(\Theta) = L_i(\delta, \theta, \beta) = 0 \quad , i = 1, 2, 3 \qquad (86)$$

where the vector Θ represents the triplet of the Weibull parameters (location, scale and shape). In this case the Newton-Raphson solution for these equations takes the form:

$$\Theta^{(k+1)} = \Theta^{(k)} + \left[L'(\Theta^{k})\right]^{-1} L(\Theta^{(k)}) \quad , k = 0, 1, ... \qquad (87)$$

where $L'(\Theta)$ denotes the Gateaux derivative of L, where L is Gateaux differentiable at Θ if $\exists$ a linear operator A $\ni$:

$$\lim_{t\to 0} \frac{\| L(\Theta + th) - L(\Theta) - tAh \|}{t} = 0 \;\; \forall h \in \mathcal{R}^{(3)} \qquad (88)$$

This method has a quadratic convergence properties, however it suffers from the pitfall of failure to converge if the initial guess $\Theta^{(0)}$ is far away from the solution Θ_s.

Several different modifications were introduced to overcome that problem. Among these methods are the norm reducing method where the derivative is multiplied by a factor such that the norm will be non-decreasing as the iterations progress. Another method is to ensure that the derivative is non-singular by adding a constant to its diagonal elements such that the new matrix is non-singular when the derivative is singular. A third method is by occasionally computing the derivative. A more detailed discussion of such methods is due to Ortega (1970).

The difficulty of such basic methods, is in the need to compute 3 components of L and 9 entries of L'. Several other modifications are introduced by Powell (1970) to alleviate such a problem by avoiding the direct computation of L' through replacing it by the difference approximations. Harter and Moore in their 1965 paper solved the system of the nonlinear equations for joint maximum likelihood estimation from complete and censored samples of the three parameter Weibull *(also of the three parameter Gamma).* The proposed iterative procedure was applied to both general case as well as cases when any one or any two of the three parameters were known. The iterative scheme used here was proposed by Powell (1970) where the derivative was not just scaled by a small factor but by introducing a negative multiple of the gradient of $L(\Theta)$ such that the direction for the correction in the different iterations

will be sensible as the Jacobian is almost singular.

The method can be applied in two cases: when the first derivative L' is given or when it is numerically approximated. Since in our case, the functional form for the derivative is not complicated, the approach when the Jacobian is given is chosen to be used.

Methodology

The technique is basically a modification of Levenberg/Marquardt idea for the classical Newton-Raphson iterative scheme for the solution of a nonlinear system of equations through the usage of:

(1) A *negative multiple* of the gradient of $L(\Theta)$ to avoid the near singularity of the Jacobian matrix.

(2) A flexible choice of the difference between $\Theta^{(k+1)}$ and $\Theta^{(k)}$ in each step is used to decrease the number of iterations depending on the increase or decrease of $L(\Theta)$.

The running time of the algorithm depends, in general, on the number of equations, the function behavior of $L(\Theta)$, the initial or the starting point $\Theta^{(0)}$, and the accuracy required in terms of the step difference and the norm.

An accuracy of .01 was used for the absolute difference between two successive $\Theta's$ while the Euclidean norm accuracy was relaxed since the MISE criteria is to be used latter for the comparison and the interest was in the convergence of the Θ

parameter mainly.

The algorithm did not converge in a few cases (24 cases) which were excluded from the Monte Carlo results. This happened because the method was searching for a zero of the system of nonlinear equations $L(\Theta)=0$ by minimizing the quadratic form $L^T(\Theta)\,L(\Theta)$ or the sum of squares of the maximum likelihood equations. In which case the minimum would not give a zero of the system.

The initial guess, is chosen to be the same for all of the different Monte Carlo samples of size 1000.

It was proved by Powell in 1970 that the iterations stops due to one of the mentioned stopping rules or otherwise the solution converges to a solution Θ_s providing that the Jacobian matrices are bounded and $L(\Theta^0)$ is finite. Powell also proved that the algorithm will stop after a finite number of iterations by one of the two stopping rules providing that $L_i(\Theta)$ is of continuous, bounded first derivatives.

Stopping Criterion

In addition, the technique introduces two stopping criterion:

First is step length in two successive iterations which is taken as .01.

Second is the maximum number of iterations which is taken as 1000.

Results

The results from the previous application are shown on tables 5 to table 17 at the end of chapter IV where cases of shape parameter 1, 2, 3 and 4 for sample sizes 10, 20, and 30 with location 10.0 and scale 5.0 are given. The tables show the sample used for each case. The integrated square error (ISE) and the function norm were also given as measures for the closeness and accuracy of the nonlinear solution. The mean integrated square error from the Monte Carlo experiment are shown at the end of the next chapter where it will be compared with the results from the minimum distance estimation technique.

VI. Minimum Distance Estimation

Introduction

Minimum distance estimation (MDE) was proposed by Wolfowitz (Wolfowitz, 1950). Parr and Schucany demonstrated the robustness of MDE in predicting the location of symmetric distributions (Parr and Schucany, 1980). Hobbs, Moore, and James (Hobbs and others, 1984) used MDE to find the location of the gamma distribution. Similarly, Hobbs, Moore, and Miller (Hobbs and others, 1985) used MDE to estimate the location of the Weibull. In recent research (Gallagher and Moore, 1989) the previous work was extended by applying MDE to all the distribution parameters and by testing the robustness of MDE.

MDE selects as estimates those p.d.f parameters which minimize the discrepancy between the sample data and the estimated distribution. The distance measures, which are minimized are " Goodness of fit statistics" (g.o.f).

The MDE has the following characterization and properties:

1. Not susceptible to outliers (Parr and Schucany, 1980).

2. Statistically consistent (Wolfowitz, 1957).

3. Easily applied to all the parameters (Parr and Schucany, 1980).

A series of logical candidates for the distance estimation task is studied by Fuchs (Fuchs, 1984).

This series includes:

- General exponential power distribution.

- Generalized beta distribution.

- Generalized gamma distribution.

- Generalized t distribution.

- R-S distribution: which was originally developed to generate random variates (Ramberg and Schmeister 1979). It is a generalization of Tukey's lambda function and can be used to model a wide variety of data.

The probability density function of the R-S distribution is given in terms of the percentile function, R(p)

$$f(x;p,a,b,c,d) = f(R(p)) = (cp^{c-1} + d(1-p)^{d-1})/b \tag{89}$$

$$R(p) = a + (p^c - (1-p)^d)/b \tag{90}$$

where $-\infty < a \leq x < \infty$; $-\infty < a, b, c, d < \infty$, $0 \leq p \leq 1$

- Generalized life model: developed by Moore and Bilikan, which includes the Weibull and the Raleigh distribution as a special case. The p.d.f is given by:

$$f\left(x; a, b, g(x)\right) = bg'(x)\left(g(x)\right)^{b-1} exp\left(-\left(g(x)\right)^{b}/a\right)/a \tag{91}$$

where $g(x) \in R^1$, $\lim_{x \to 0^+} g(x) = 0$, $\lim_{x \to \infty} g(x) = \infty$

and $g(x)$ is strictly increasing ,$0 < x$, $a, b < \infty$

Minimum Distance Estimation For The Three Parameter Weibull Distribution

The 3-parameter Weibull density function is given by:

$$f(x) = \frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{\beta-1} exp\left[-\left(\frac{x-\delta}{\theta}\right)^{\beta}\right] \quad ,\delta \leq x\,, \theta, \beta > 0 \qquad (92)$$

with expected value

$$E(x) = \delta + \theta\Gamma\left(\frac{\beta+1}{\beta}\right) \qquad (93)$$

and with variance

$$V(x) = \theta^2\left[\Gamma\left(\frac{\beta+2}{\beta}\right) - \Gamma^2\left(\frac{\beta+1}{\beta}\right)\right] \qquad (94)$$

where Γ denotes the gamma function.

and C.D.F

$$F(x) = 1 - e^{-\left(\frac{x-\delta}{\theta}\right)^{\beta}} \qquad (95)$$

It is required to estimate δ , θ , β such that a g.o.f statistic is minimized using a non-parametric estimator $\hat{f}(x)$ for f(x). The kernel estimator discussed in chapter

III is used with a Gaussian kernel which is defined as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{96}$$

$$= \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\, exp\left(\frac{x - X_i}{h}\right)^2 \tag{97}$$

The C.D.F of this kernel density $\hat{F}(x)$ is given as:

$$\hat{F}(x) = \int_{-\infty}^{x} \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\, exp\left(\frac{x - X_i}{h}\right)^2 dx \tag{98}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}h}\, exp\left(\frac{x - X_i}{h}\right)^2 dx \tag{99}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{x - X_i}{h}\right) \tag{100}$$

where $\Phi(x)$ denotes the C.D.F for a standard normal random variable.

The Cramer von Mises statistic W^2 is used. This g.o.f. statistic is defined as:

$$W^2 = n\int_{-\infty}^{\infty} \left[\hat{F}(x) - F_o(x)\right]^2 dF_o(x) \tag{101}$$

or the computational formula:

$$W^2 = \sum_{j=1}^{n}\left[F(x_j) - \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{x - X_i}{h}\right)\right]^2 + \frac{1}{12n} \tag{102}$$

As it was noted early, the optimal value of the window width h (in the MISE sense) depends on the choice of the kernel K, the underlying unknown density f(x) and the sample size i.e

$$h_{opt} = f_1(K).f_2(f(x)).f_3(n) \tag{103}$$

A reasonable approximation for this optimal value for a normal sample is $h = kn^{-\frac{1}{5}}$ where k is a real constant (see equation 38). Although this approximation simplifies the optimal expression for the window width and works fine with the normal distribution, it is not as good for other distributions. This leads to the idea of introducing the underlying density in another approximating expression for that h. The explicit expression for h_{opt} is given as:

$$h_{opt} = m_2^{-2/5} \left\{ \int K^2(t)dt \right\}^{1/5} \left\{ \int f''(x)^2 \, dx \right\}^{-1/5} n^{-1/5} \tag{104}$$

where:

m_2 denotes the kernel second moment.

In case of a Gaussian kernel:

$$\begin{aligned} m_2 &= \int t^2 K(t)\, dt \\ &= V(t) \\ &= 1 \end{aligned} \tag{105}$$

also, $\int K^2(t)\,dt$ is simply equal to $\frac{1}{2\sqrt{\pi}}$

Now,let

$$S_i(x) = exp\left(\frac{-1}{2}\left(\frac{x - X_i}{h}\right)^2\right) \tag{106}$$

and

$$l_i(x) = -\left(\frac{x - X_i}{h^2}\right) \tag{107}$$

Hence, $\hat{f}(x)$ can be written as

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{n} S_i(x) \tag{108}$$

and

$$\hat{f}'(x) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{n} S_i'(x) \tag{109}$$

but

$$\begin{aligned} S_i'(x) &= exp\left(\frac{-1}{2}\left(\frac{x - X_i}{h}\right)^2\right)\left[-\left(\frac{x - X_i}{h^2}\right)\right] \\ &= S_i(x)l_i(x) \end{aligned} \tag{110}$$

also,

$$l_i'(x) = -\frac{1}{h^2} \tag{111}$$

$$\hat{f}'(x) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{n} S_i(x)l_i(x) \tag{112}$$

$$\hat{f}''(x) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{n}\left[l_i^2(x)S_i(x) - \frac{S_i(x)}{h^2}\right] \tag{113}$$

$$\hat{f}''^2(x) = \frac{1}{2\pi n^2 h^2}\left\{\sum_{i=1}^{n} S_i(x)\left[l_i^2(x) - \frac{1}{h^2}\right]\right\}^2 \tag{114}$$

since

$$\left[\sum_{k=1}^{n}\zeta_k\right]^2 = \sum_{i,j=1}^{n}\zeta_i\zeta_j \tag{115}$$

thus

$$\hat{f}''^2(x) = \frac{1}{2\pi n^2 h^2}\sum_{i,j=1}^{r}\left\{S_i(x)S_j(x)\left[l_i(x)^2 - \frac{1}{h^2}\right]\left[l_j(x)^2 - \frac{1}{h^2}\right]\right\} \tag{116}$$

$$S_i(x)S_j(x) = exp\left[-\frac{1}{2}\left(\frac{x-X_i}{h}\right)^2 - \frac{1}{2}\left(\frac{x-X_j}{h}\right)^2\right]$$

$$= exp\left[-\frac{1}{2h^2}\left(2x^2 - 2x\left(X_i + X_j\right) + X_i^2 + X_j^2\right)\right]$$

$$= exp\left\{-\frac{1}{4h^2}\left[x^2 - x\left(X_i + X_j\right) + \left(\frac{X_i + X_j}{2}\right)^2 - \left(\frac{X_i + X_j}{2}\right)^2 + \frac{X_i^2 + X_j^2}{2}\right]\right\}$$

$$= exp\left[-\frac{1}{16h^2}\left(X_i - X_j\right)^2\right] exp\left[-\frac{1}{4h^2}\left(x - \frac{X_i + X_j}{2}\right)^2\right]$$

$$= \quad exp\left[-\frac{1}{16h^2}(X_i - X_j)^2\right] g_{ij}(x).2\sqrt{\pi}h \tag{117}$$

where $g_{ij}(x)$ is a normal density distribution with mean $\frac{X_i+X_j}{2}$, and variance $2h^2$.

Now, let $l_i(x)$, $l_j(x)$ be written as l_i, l_j for the simplicity of the notations.

$$\left(l_i^2(x) - \frac{1}{h^2}\right)\left(l_j^2(x) - \frac{1}{h^2}\right) \quad = \quad l_i^2 l_j^2 + \frac{1}{h^4} - \frac{l_i^2 + l_j^2}{h^4}$$

$$= \quad \frac{x^4 - 2x^3(X_i + X_j) + x^2(X_i^2 + X_j^2 + 4X_iX_j) - 2xX_iX_j(X_i + X_j)}{h^8}$$

$$- \frac{2x^2 - 2x(X_i + X_j) + X_i^2 + X_j^2}{h^6} + \frac{1}{h^4} + \frac{X_i^2 X_j^2}{h^8}$$

$$= \quad \frac{x^4}{h^8} - \frac{2x^3(X_i + X_j)}{h^8} + x^2\left[\frac{X_i^2 + X_j^2 + 4X_iX_j}{h^8} + \frac{2}{h^6}\right]$$

$$- 2x\left[\frac{X_iX_j(X_i + X_j)}{h^8} + \frac{X_i^2 + X_j^2}{h^6}\right] + \frac{X_i^2X_j^2}{h^8} + \frac{X_i^2 + X_j^2}{h^6} + \frac{1}{h^4}$$

thus,

$$\hat{f}''^2(x) \quad = \quad \sum_{i,j} \frac{e^{\frac{-(X_i - X_j)^2}{16h^2}}}{\sqrt{\pi}n^2} g_{ij}(x)\left\{\frac{x^4}{h^9} - \frac{2x^3(X_i + X_j)}{h^9} + x^2\left[\frac{X_i^2 + X_j^2 + 4X_iX_j}{h^9} + \frac{2}{h^7}\right]\right.$$

$$\left. - 2x\left[\frac{X_iX_j(X_i + X_j)}{h^9} + \frac{X_i^2 + X_j^2}{h^7}\right] + \frac{X_i^2X_j^2}{h^9} + \frac{X_i^2 + X_j^2}{h^7} + \frac{1}{h^5}\right\}$$

$$\int_x \hat{f}''^2(x)dx = \sum_{i,j} \frac{e^{\frac{-(X_i-X_j)^2}{16h^2}}}{\sqrt{\pi}n^2} \left\{ \frac{E(x^4)}{h^9} - \frac{2E(x^3)(X_i+X_j)}{h^9} \right.$$

$$+ E(x^2)\left[\frac{X_i^2+X_j^2+4X_iX_j}{h^9}+\frac{2}{h^7}\right]$$

$$\left. - 2E(x)\left[\frac{X_iX_j(X_i+X_j)}{h^9}+\frac{X_i^2+X_j^2}{h^7}\right]+\frac{X_i^2X_j^2}{h^9}+\frac{X_i^2+X_j^2}{h^7}+\frac{1}{h^5}\right\}$$

where

$$E(x^4) = \left(\frac{X_i+X_j}{2}\right)^4 + 12\left(\frac{X_i+X_j}{2}\right)^2 h^2 + 12h^4 \tag{118}$$

$$E(x^3) = \frac{X_i+X_j}{2}\left[\left(\frac{X_i+X_j}{2}\right)^2 + 6h^2\right] \tag{119}$$

$$E(x^2) = \left(\frac{X_i+X_j}{2}\right)^2 + 2h^2 \tag{120}$$

$$E(x) = \frac{X_i+X_j}{2} \tag{121}$$

Hence

$$\int_x \hat{f}''^2(x)\,dx = \sum_{i,j} \frac{e^{\frac{-(X_i-X_j)^2}{16h^2}}}{\sqrt{\pi}n^2} \left\{ \frac{1}{h^9}\left[\left(\frac{X_i+X_j}{2}\right)^4 + 12h^2\left(\frac{X_i+X_j}{2}\right)^2 + 12h^4\right]\right.$$

$$- \frac{1}{h^9} \left[\frac{(X_i + X_j)^2}{2} \left(\frac{X_i + X_j}{2} \right)^2 + 6h^2 \right]$$

$$+ \left[\left(\frac{X_i + X_j}{2} \right)^2 + 2h^2 \right] \left[\frac{X_i^2 + X_j^2 + 4X_iX_j}{h^9} + \frac{2}{h^7} \right]$$

$$- (X_i + X_j) \left[\frac{X_iX_j (X_i + X_j)}{h^9} + \frac{X_i^2 + X_j^2}{h^7} \right]$$

$$+ \frac{X_i^2 X_j^2}{h^9} + \frac{X_i^2 + X_j^2}{h^7} + \frac{1}{h^5} \Bigg\} \tag{122}$$

On substituting this previous integral for the integral of the density squared in the expression for the optimal h, h will be possibly written as:

$$h_{opt} = \Upsilon(h) \tag{123}$$

or equivalently as:

$$\Upsilon_1(h) = h_{opt} - \Upsilon(h) = 0 \tag{124}$$

which can be solved by one of the generalization methods for the solution of one equation in one unknown, such as Newton's method, secant method, Steffenson's method or any of their variations.

The Newton's method has the form:

$$h^{k+1} = h^k - \left[\Upsilon_1' \left(h^k \right) \right]^{-1} \Upsilon_1 \left(h^k \right) \tag{125}$$

which gives a quadratic convergence i.e

$$||h^{k+1} - h^*|| \leq c||h^k - h^*|| \tag{126}$$

for a sufficiently close h^k , h^*

An alternative for computing the window width which is more efficient computationally and gives a good improvement in this application is to choose an empirical h which equals $sn^{-1/5}$ where s represents the sample standard deviation. This suggested h showed MISE which is close enough to the optimal theoretical and since it was simple, without a need to extensive computations and face degeneracy sometimes compared to the iterative approach.

Methodology

The Monte Carlo procedure for this application can be described in the following three steps:

Step I

- Different samples from Weibull with a given location, scale, and shape for different sample sizes are generated. The uniform random number is generated using the RNUN routine from the IMSL.

- The Weibull deviates are generated using the inverse C.D.F technique.

Step II

- The MLE estimators for the 3-parameters are computed as discussed earlier.

- The CvM statistic is computed for the estimated density with MLE for the parameters.

<u>**Step III**</u>

- Minimizing the CvM statistic with Θ as the decision vector and with the given constraints on the values of the parameters.

- The non-linear program is solved using quasi Newton method.

- The new parameter estimates are compared with those of MLE, Using the ISE as a measure for the comparison.

Results

Together with the results from the previous chapter, the end result for this application is shown in tables 6. The table shows that both the MLE method and the new technique are statistically the same for shape parameter 1. However the new technique shows a significant improvement over the MLE method for shape parameters 2, 3, and 4. For shape parameter 2 the new method gives an MISE which is 5.3 times smaller than that of the MLE, while in the case of shape parameter 3 the MISE from the new technique is about 6 times smaller than that of the MLE. For shape parameter 4 a tremendous improvement is obtained, where the ratio between the MISE for MLE to that of the new tecnique is 15.9, which shows how big the improvement is due to the new technique. Table 7 to table 18 give examples

from each of the four shape parameter values chosen for the Monte Carlo. These tables give a case for each value of the shape parameter 1, 2, 3, and 4 for sample sizes 10, 20, and 30 with location 10.0 and scale 5.0. The same sample is used to iteratively solve the maximum likelihood nonlinear equations. The integrated square error (ISE), the value of the window width used, and the optimal value for the CvM statistic based on using the nonparametric density estimation approach are given. The graphs for these cases are given in figures 1 to 24 while the next table shows the resulting MISE together with its standard deviation for sample size 20 for the different parameter values for both the new proposed estimation technique concurrently with the modified nonlinear method for solving the ML equations.

Table 6. Results from M.C size 1000 for sample size 20

Weibull(loc., sca., sha.)	$MISE_{CvM}$	$MISE_{MLE}$
W(10,5,1)	.13209	.13678
	(.17234)	*(.17820)*
W(10,5,2)	.04970	.26364
	(.05061)	*(.19757)*
W(10,5,3)	.03378	.20255
	(.03385)	*(.17740)*
W(10,5,4)	.02575	.40923
	(.02551)	*(.49626)*

Table 7 to table 18 show that the choice of the h parameter varies from sample to sample and from one shape parameter to another. The tables also show variations

in the value of MISE over different shape parameters for the Weibull density. These variations in h value together with the variations in the MISE indicate that that the method used is an adaptive one in the sense that the choice of the parameter h which is data dependent varies with the variation of the distribution shape and the particular sample.

Thus the final conclusion is the minimum distance estimation method using the CvM statistic as a measure of the difference between a nonparametric estimator based on a suggested window width and a parametric density with unknown parameters gives in general a much smaller MISE value than the maximum likelihood method.

Table 7. Weibull Sample (Shape = 1.0 and Sample Size = 10)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 1.0

SAMPLE SIZE = 10

Weibull Data Values

10.009320
10.226890
10.798260
10.866060
11.054560
11.788680
14.245620
14.910420
17.955210
24.277580

	TRUE	MLE	MDCVM
LOCATION	10.0000	10.0080	7.4580
SCALE	5.0000	3.4000	6.8850
SHAPE	1.0000	0.9500	1.2990
ISE		0.1398	0.1253

Function Norm	3011.9329
Window Width	2.8563
Optimal CvM	0.0090

Parameter estimation for the three parameter
Weibull density W(10,5,1)
Sample size 10
using nonparametric modifeid MDE technique

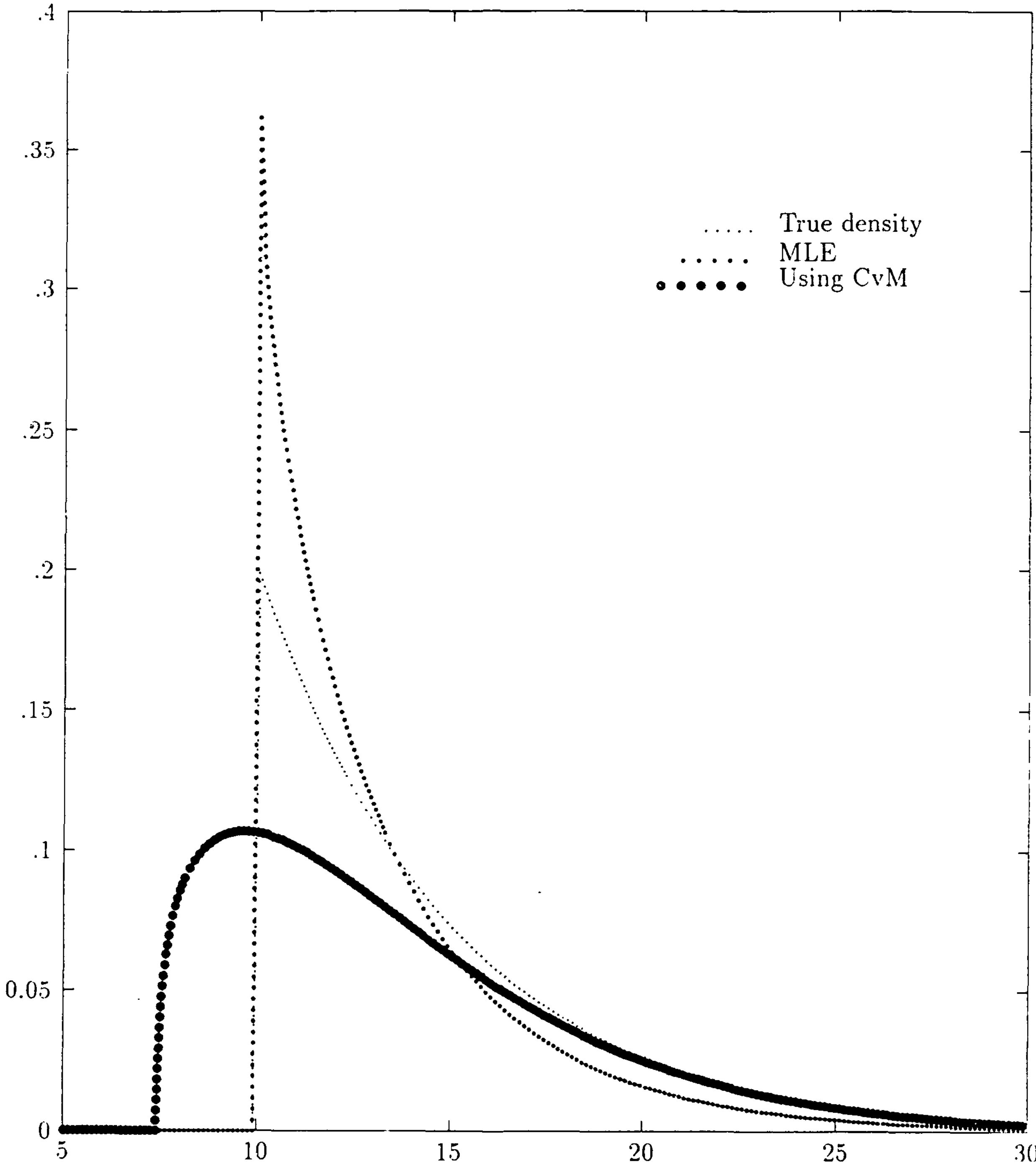

Figure 13. p.d.f for W(10,5,1) with N=10

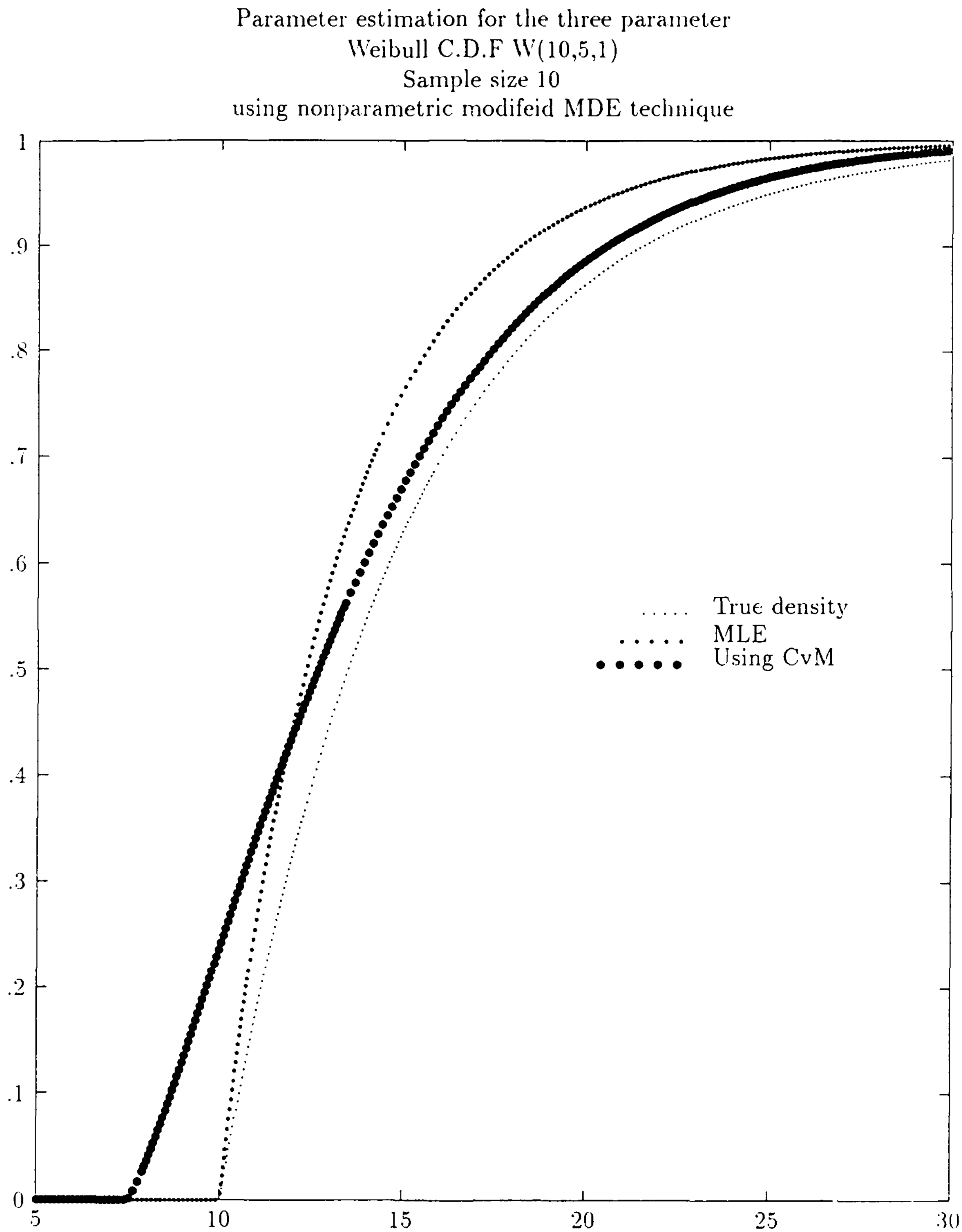

Figure 14. C.D.F. for W(10,5,1) with N=10

Table 8. Weibull Sample (Shape = 2.0 and Sample Size = 10)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 2.0

SAMPLE SIZE = 10

Weibull Data Values

10.905100
12.259120
13.099780
14.168940
14.219130
15.972290
16.971910
16.296600
16.373850
16.426060

	TRUE	MLE	MDCVM
LOCATION	10.0000	10.9041	0.4785
SCALE	5.0000	2.9963	14.9260
SHAPE	2.0000	0.9469	7.1468
ISE		0.1231	0.0221
Function Norm		3234.4231	
Window Width		1.2068	
Optimal CvM		0.0085	

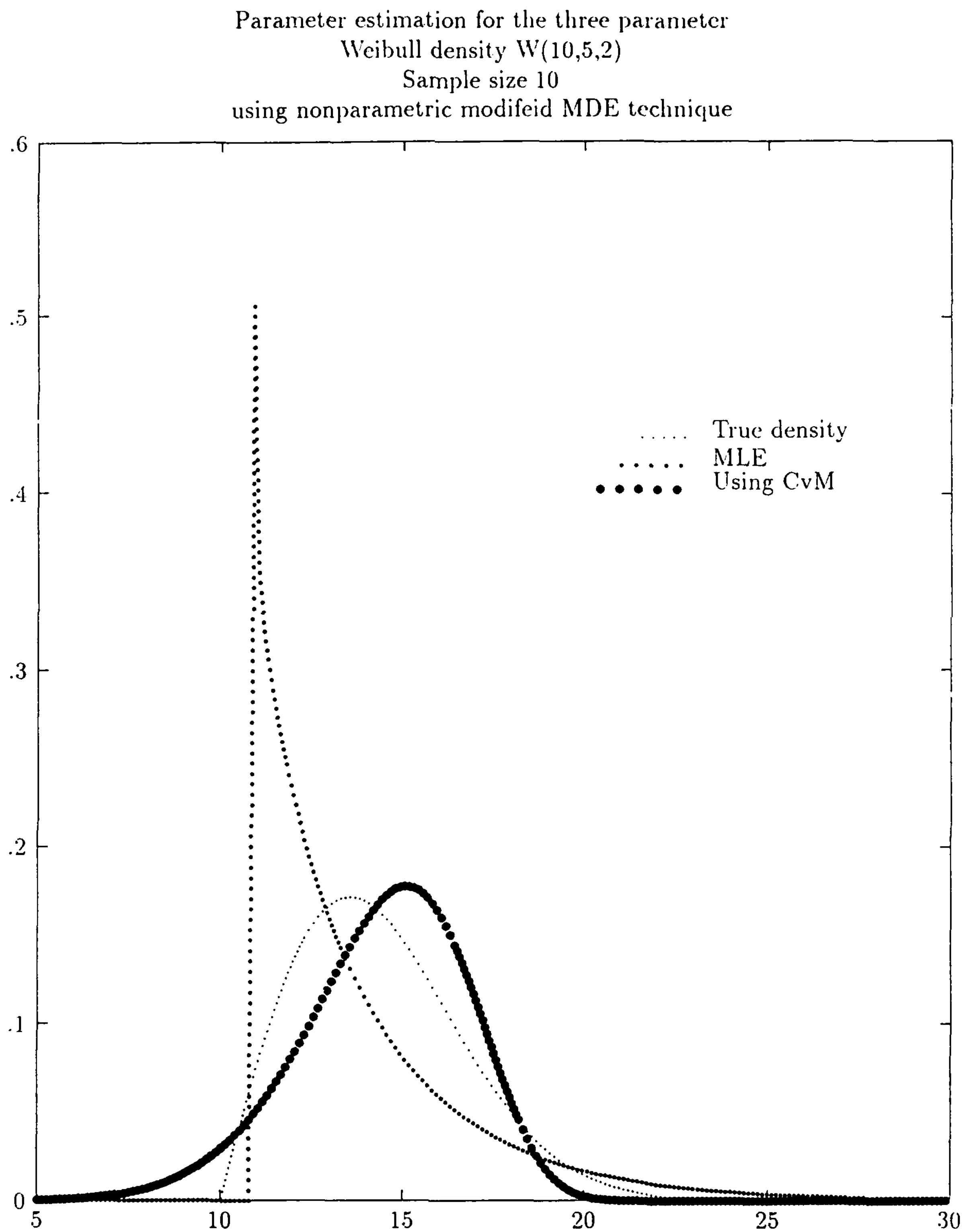

Figure 15. p.d.f for W(10,5,2) with N=10

Parameter estimation for the three parameter
Weibull C.D.F W(10,5,2)
Sample size 10
using nonparametric modifeid MDE technique

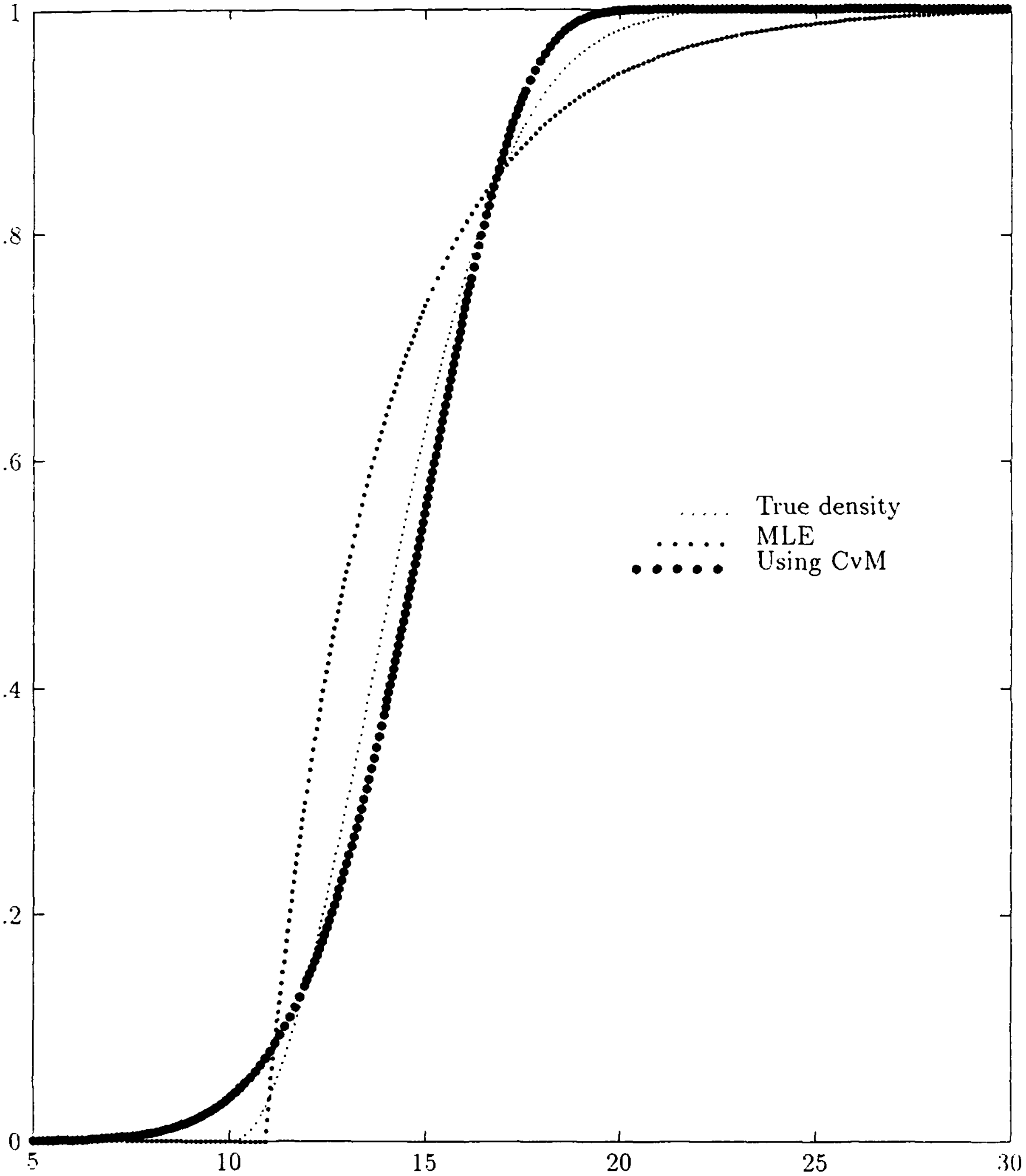

Figure 16. C.D.F. for W(10,5,2) with N=10

Table 9. Weibull Sample (Shape = 3.0 and Sample Size = 10)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 3.0

SAMPLE SIZE = 10

Weibull Data Values

11.783980
12.080590
12.377060
13.530170
13.606270
13.616520
13.776460
13.902000
15.878190
16.802490

	TRUE	MLE	MDCVM
LOCATION	10.0000	1.7830	10.6560
SCALE	5.0000	4.8683	3.5345
SHAPE	3.0000	1.1199	1.7830
ISE		0.3463	0.1040
Function Norm		13781.3096	
Window Width		0.9975	
Optimal CvM		0.0086	

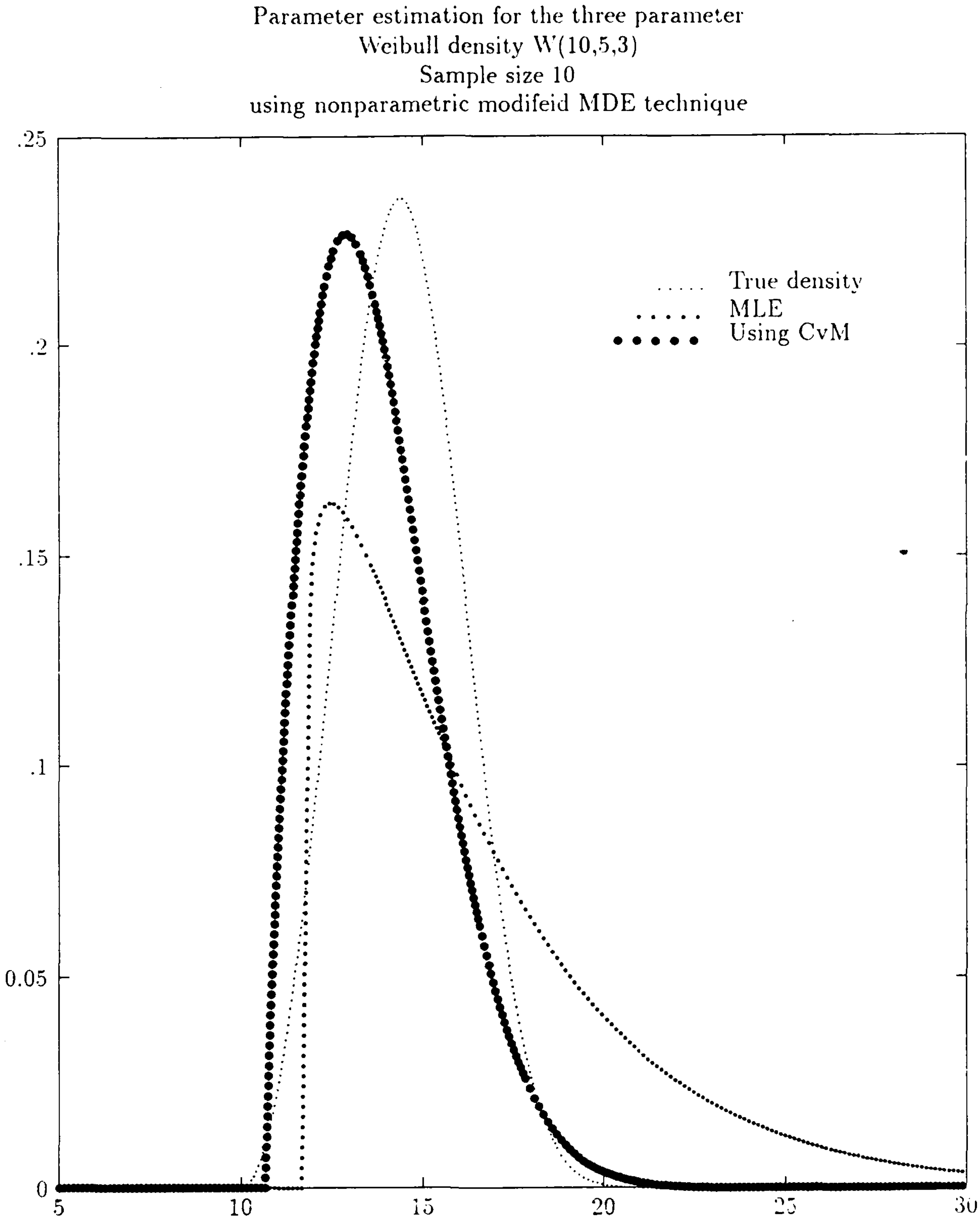

Figure 17. p.d.f for W(10,5,3) with N=10

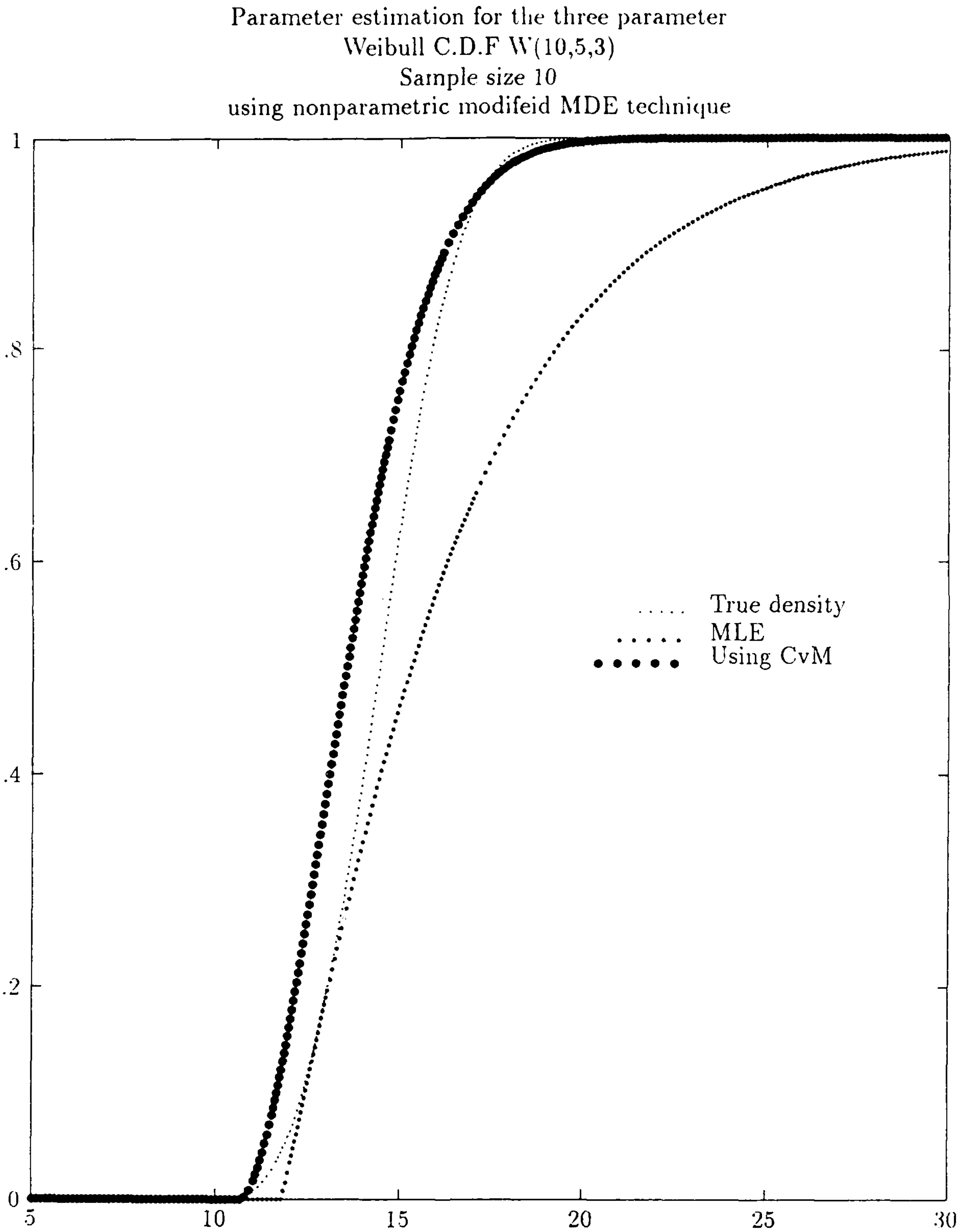

Figure 18. C.D.F. for W(10,5,3) with N=10

Table 10. Weibull Sample (Shape = 4.0 and Sample Size = 10)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 4.0

SAMPLE SIZE = 10

Weibull Data Values

11.963030
12.860350
13.219490
14.176460
14.512920
14.892670
14.919950
15.371890
15.641700
16.557470

	TRUE	MLE	MDCVM
LOCATION	10.0000	11.9620	6.1854
SCALE	5.0000	1.4014	8.8608
SHAPE	4.0000	1.6940	5.8242
ISE		0.5027	0.0134

Function Norm	478115.3125
Window Width	0.8778
Optimal CvM	0.0085

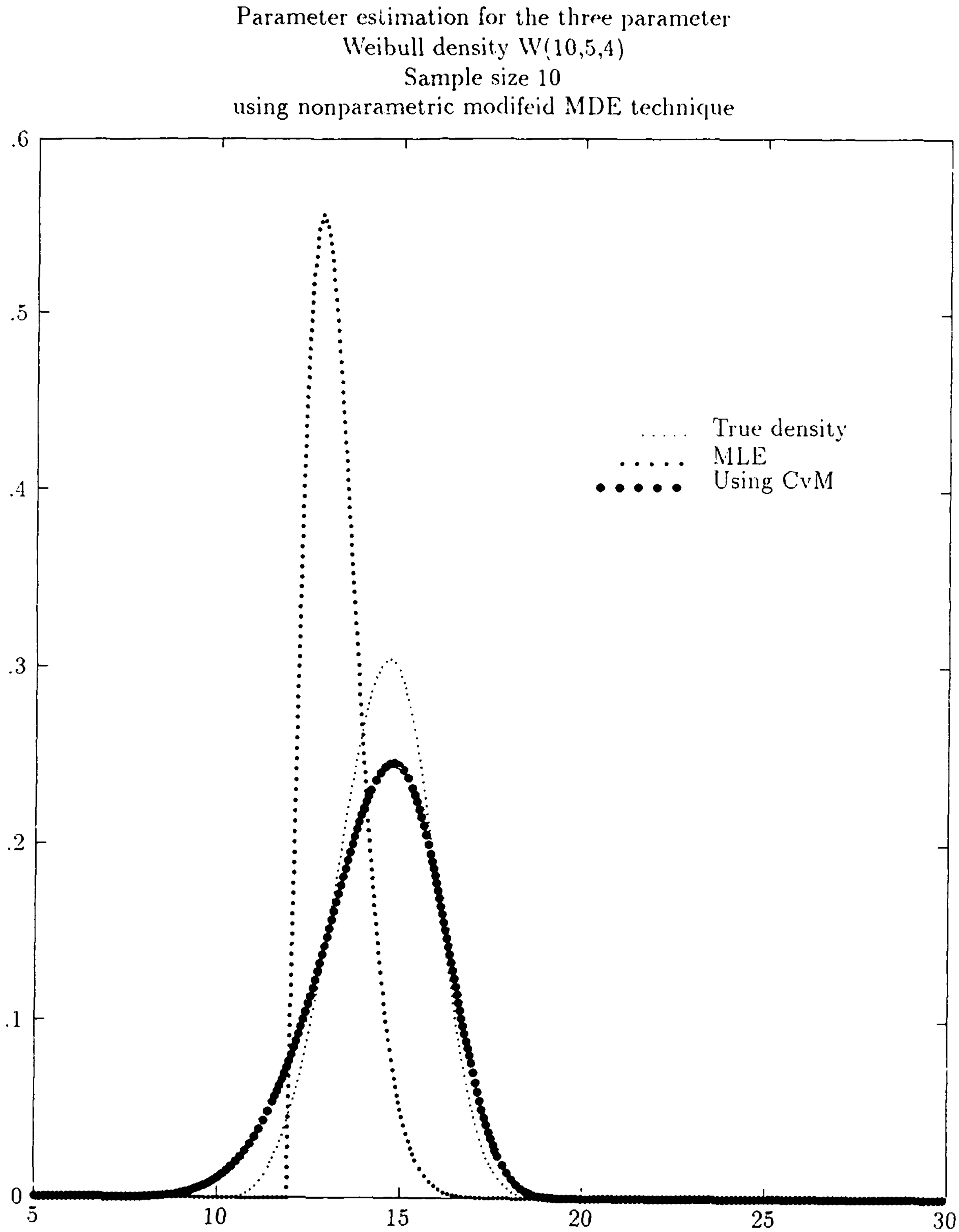

Figure 19. p.d.f for W(10,5,4) with N=10

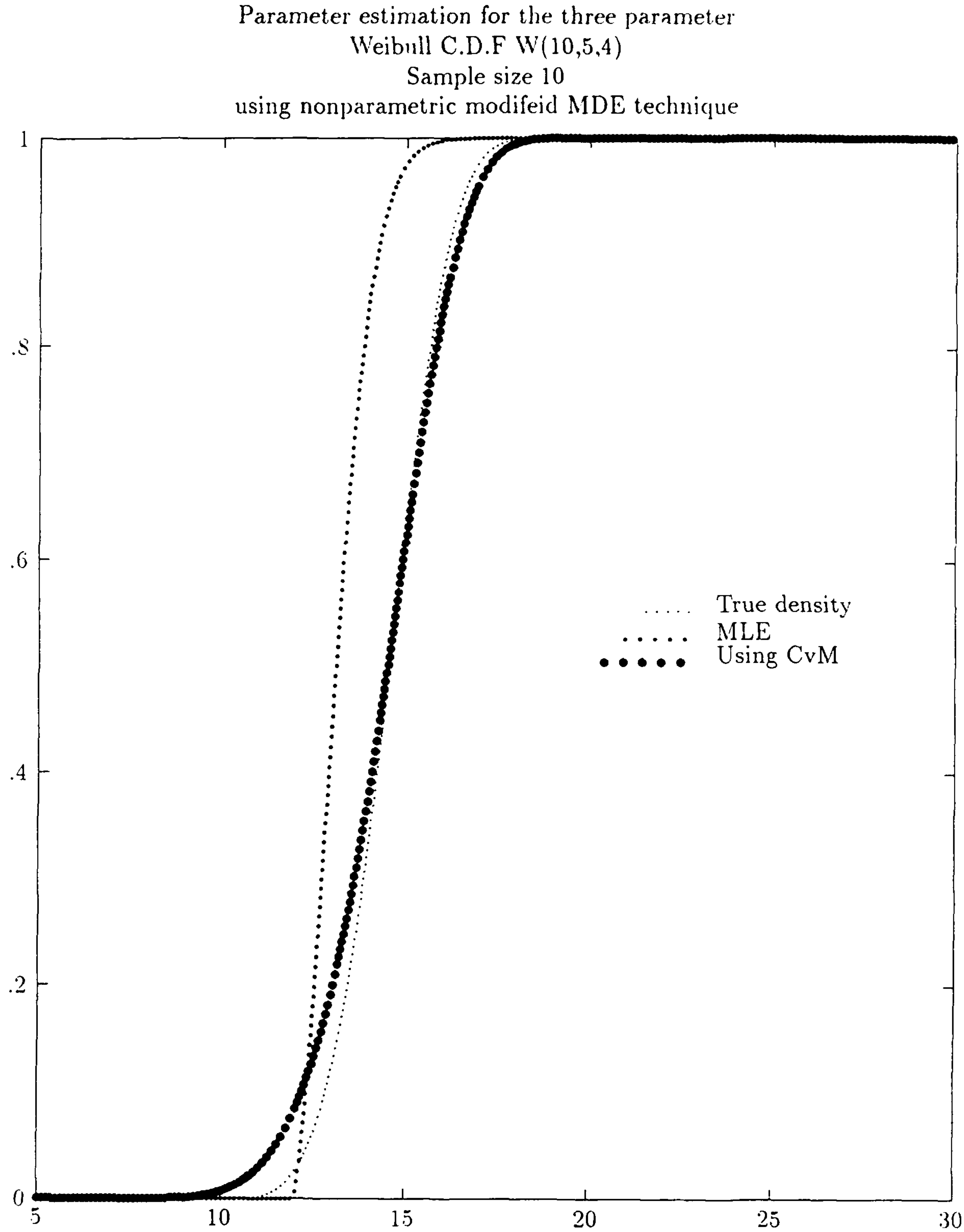

Figure 20. C.D.F. for W(10,5,4) with N=10

Table 11. Weibull Sample (Shape = 1.0 and Sample Size = 20)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 1.0

SAMPLE SIZE = 20

Weibull Data Values

10.058190	12.376410
10.227110	12.453510
10.360260	12.490680
10.423800	13.050770
10.537260	14.149840
11.759740	14.439350
11.876000	17.355110
11.892060	18.124390
11.906630	18.503469
12.154340	22.591120

	TRUE	MLE	MDCVM
LOCATION	10.0000	10.0572	8.7660
SCALE	5.0000	3.0587	5.0997
SHAPE	1.0000	0.9860	1.2873
ISE		0.2165	0.1183
Function Norm		508.4898	
Window Width		1.8328	
Optimal CvM		0.0058	

Parameter estimation for the three parameter
Weibull density W(10,5,1)
Sample size 20
using nonparametric modifeid MDE technique

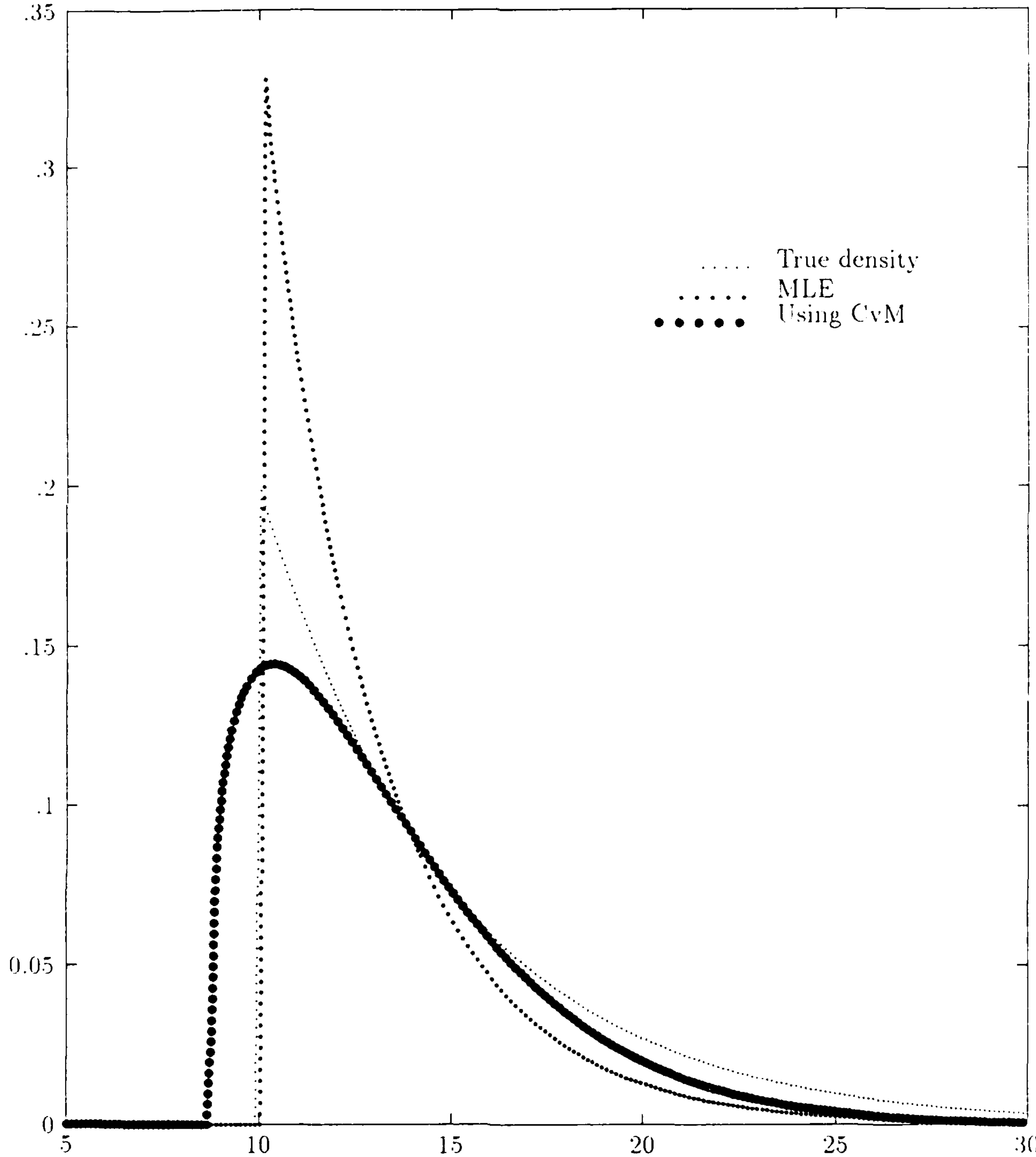

Figure 21. p.d.f for W(10,5,1) with N=20

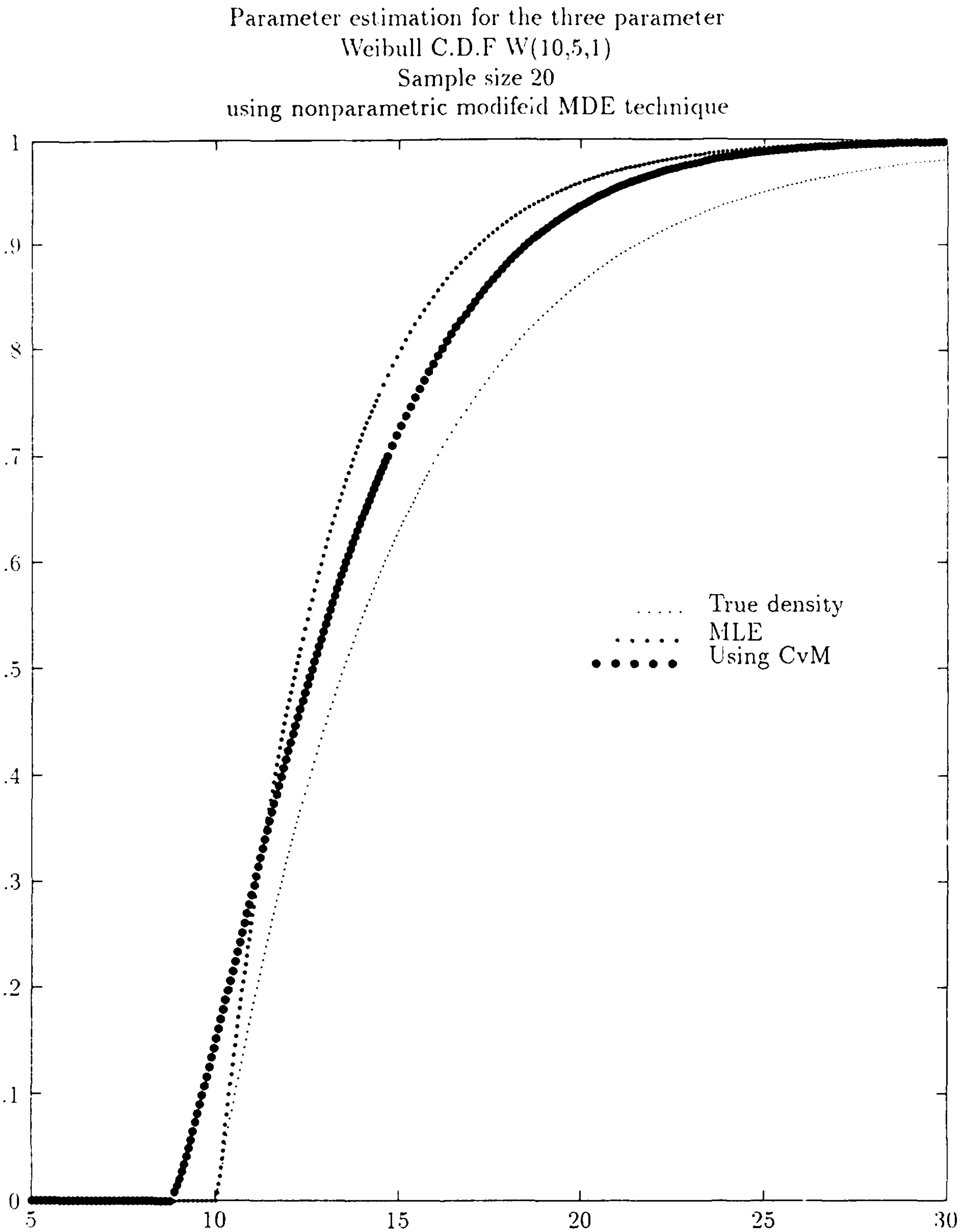

Figure 22. C.D.F. for W(10,5,1) with N=20

Table 12. Weibull Sample (Shape = 2.0 and Sample Size = 20)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 2.0

SAMPLE SIZE = 20

Weibull Data Values

10.539380	13.447030
11.065610	13.502510
11.342130	13.528930
11.455670	13.905620
11.638990	14.555130
12.966260	14.711340
13.062680	16.064280
13.075760	16.373541
13.087580	16.520531
13.282030	17.934460

	TRUE	MLE	MDCVM
LOCATION	10.0000	10.5384	9.0263
SCALE	5.0000	2.1504	5.2000
SHAPE	2.0000	1.0423	2.2099
ISE		0.4962	0.0737

Function Norm 1238.4871

Window Width 1.0842

Optimal CvM 0.0050

Parameter estimation for the three parameter
Weibull density W(10,5,2)
Sample size 20
using nonparametric modifeid MDE technique

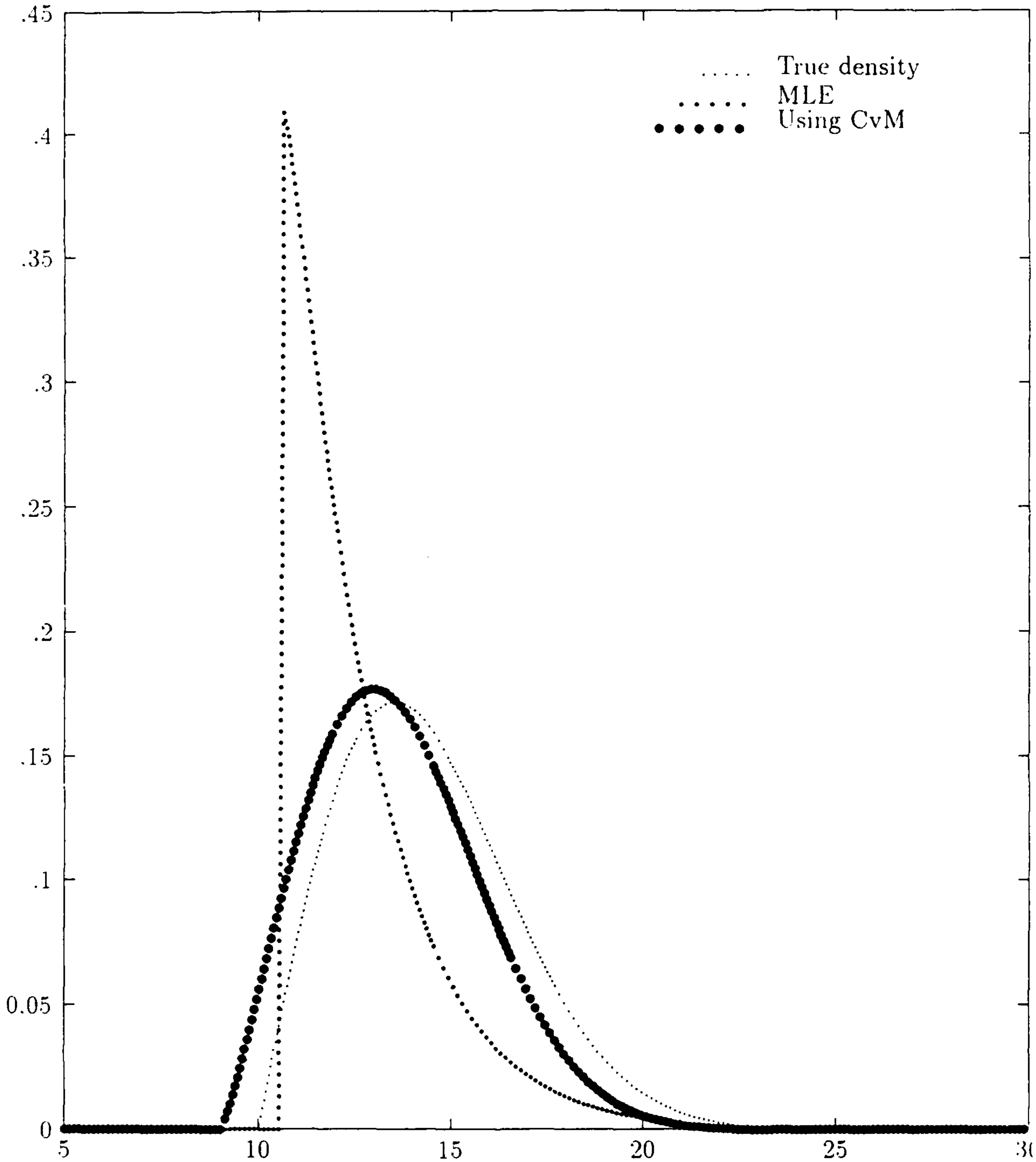

Figure 23. p.d.f for W(10,5,2) with N=20

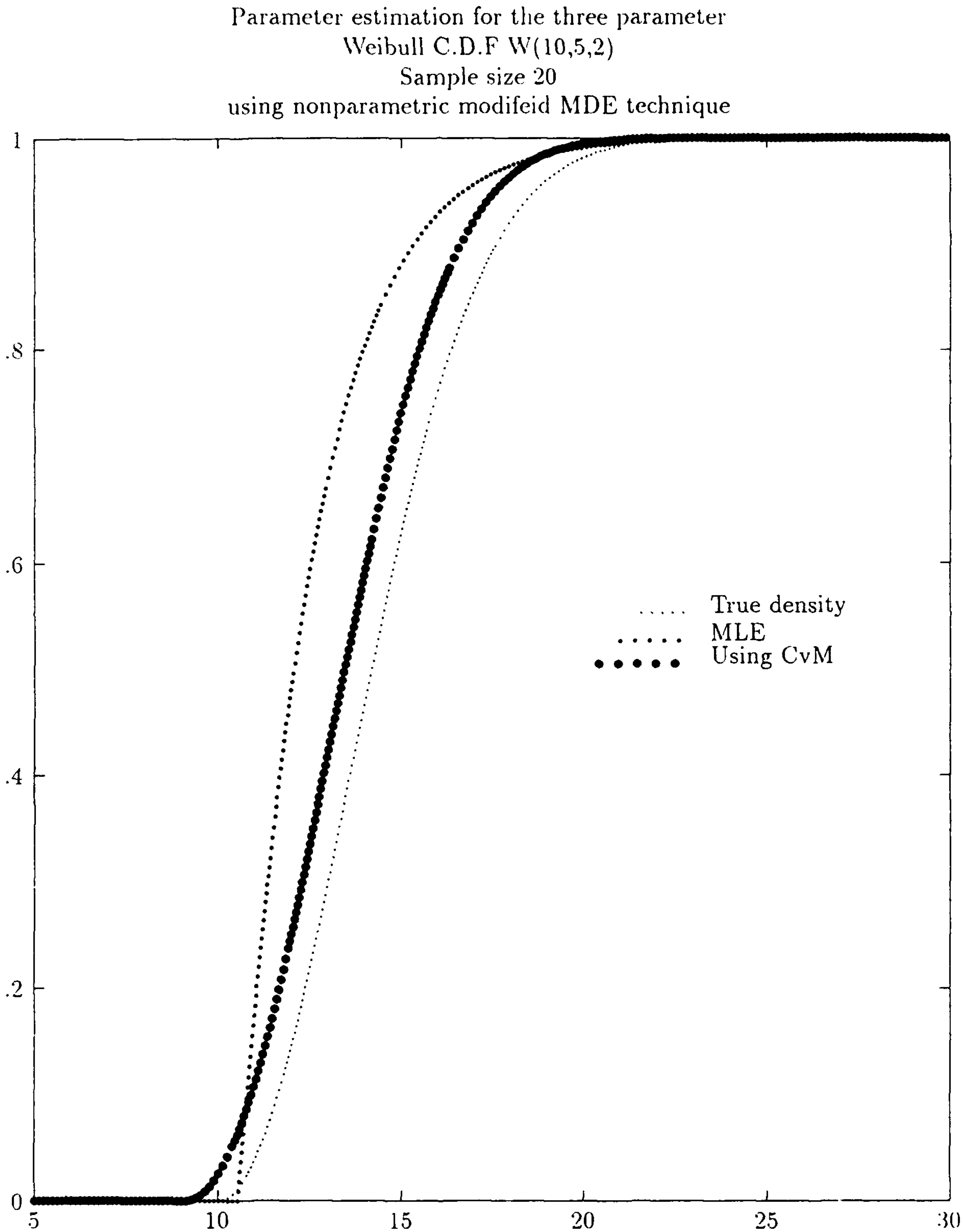

Figure 24. C.D.F. for W(10,5,2) with N=20

Table 13. Weibull Sample (Shape = 3.0 and Sample Size = 20)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 3.0

SAMPLE SIZE = 20

Weibull Data Values

11.133070	13.902000
11.783980	13.943750
12.080590	13.963560
12.196340	14.240820
12.377060	14.698840
13.530170	14.805660
13.606270	15.686470
13.616520	15.878190
13.625780	15.968230
13.776460	16.802490

	TRUE	MLE	MDCVM
LOCATION	10.0000	11.1321	8.9949
SCALE	5.0000	3.0818	5.4584
SHAPE	3.0000	1.1842	3.2065
ISE		0.1308	0.0552
Function Norm		31529.4902	
Window Width		0.8209	
Optimal CvM		0.0047	

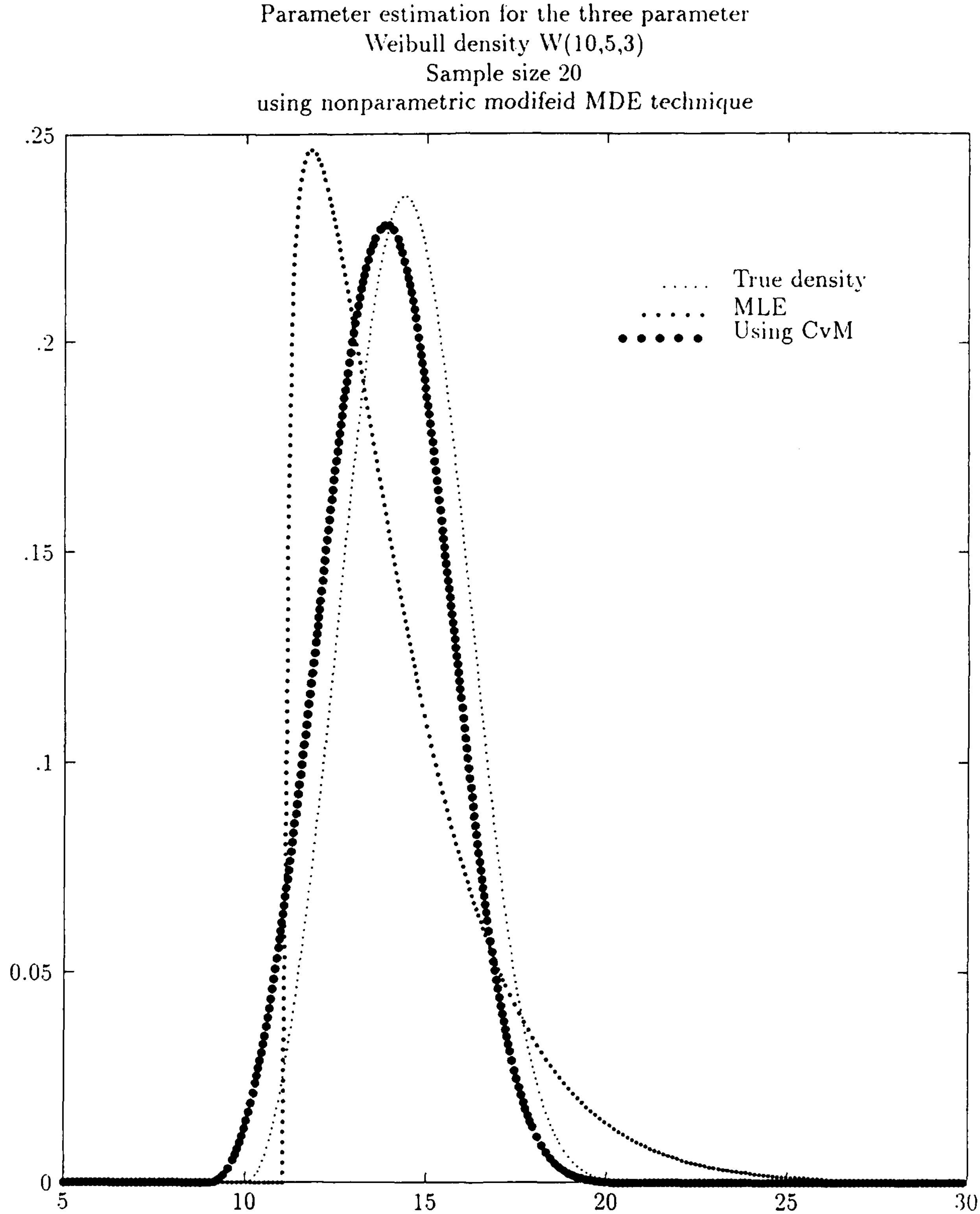

Figure 25. p.d.f for W(10,5,3) with N=20

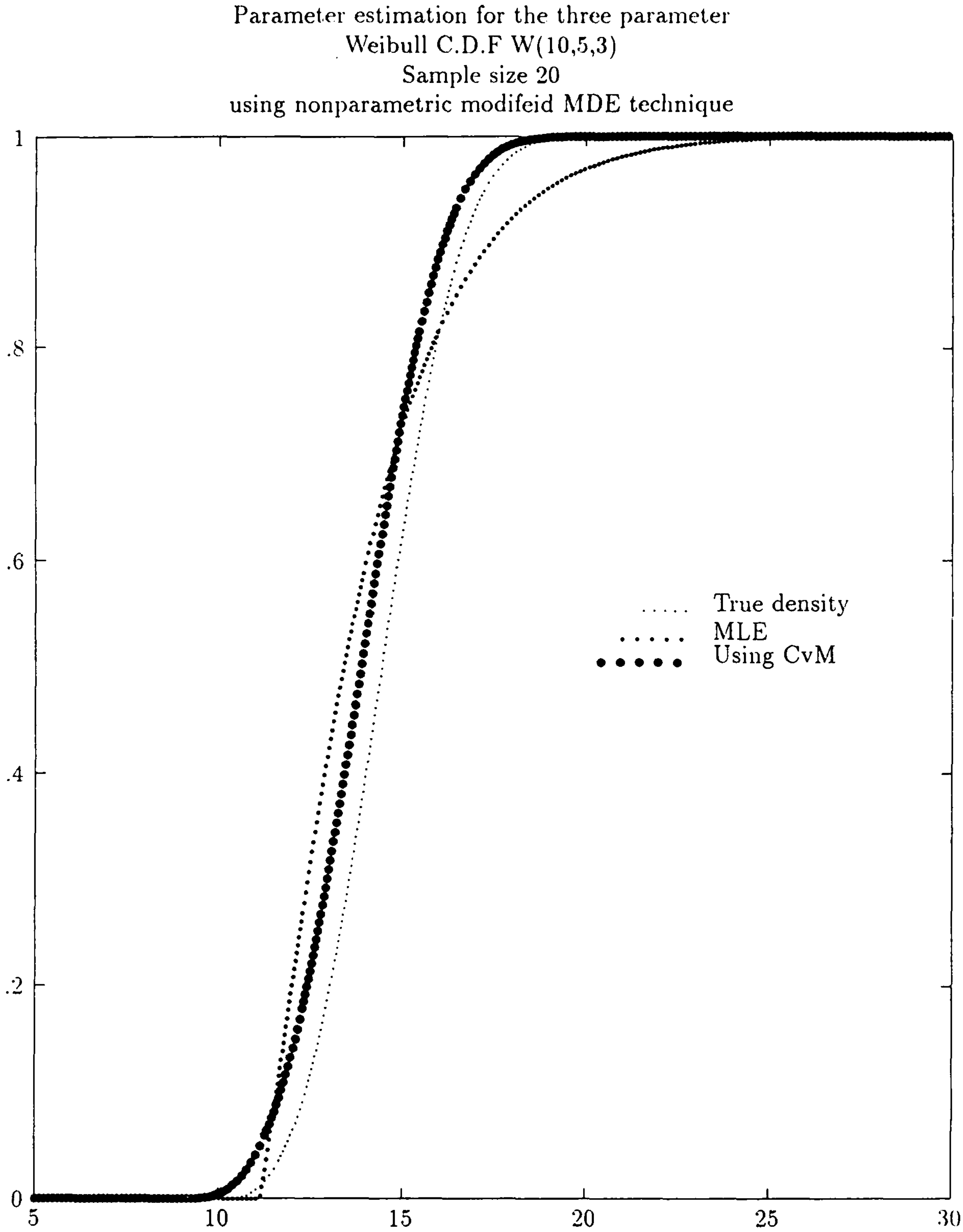

Figure 26. C.D.F. for W(10,5,3) with N=20

Table 14. Weibull Sample (Shape = 4.0 and Sample Size = 20)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 4.0

SAMPLE SIZE = 20

Weibull Data Values

11.963030	14.216850
12.212220	14.512920
12.269390	14.717380
12.331440	14.892670
12.860350	14.919950
13.062320	15.143880
13.219490	15.371890
13.540850	15.641700
14.020930	16.008890
14.176460	16.557470

	TRUE	MLE	MDCVM
LOCATION	10.0000	11.9620	8.0986
SCALE	5.0000	3.4861	6.5663
SHAPE	4.0000	1.0087	4.1925
ISE		0.1814	0.0497
Function Norm		3.8416	
Window Width		0.7450	
Optimal CvM		0.0051	

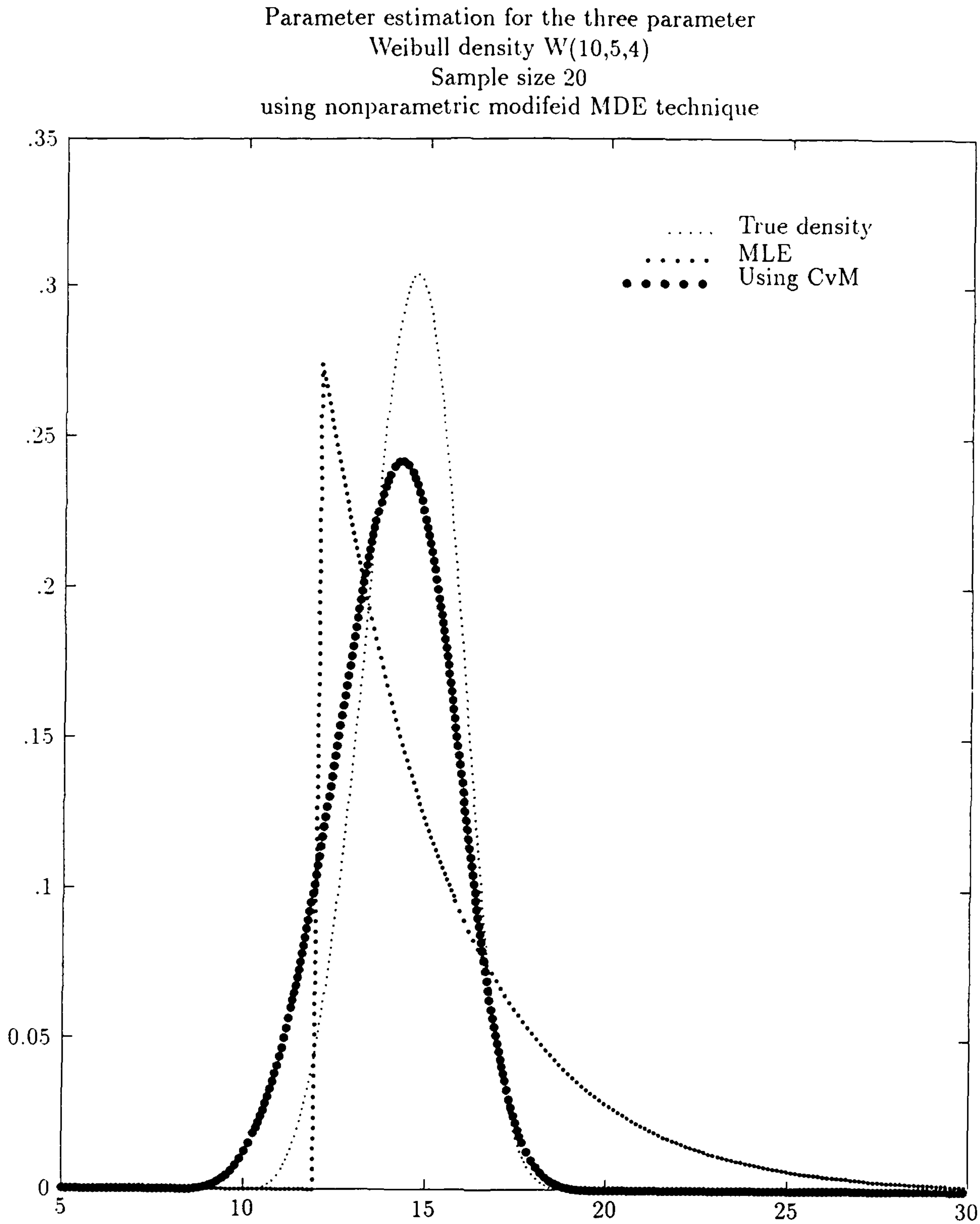

Figure 27. p.d.f for W(10,5,4) with N=20

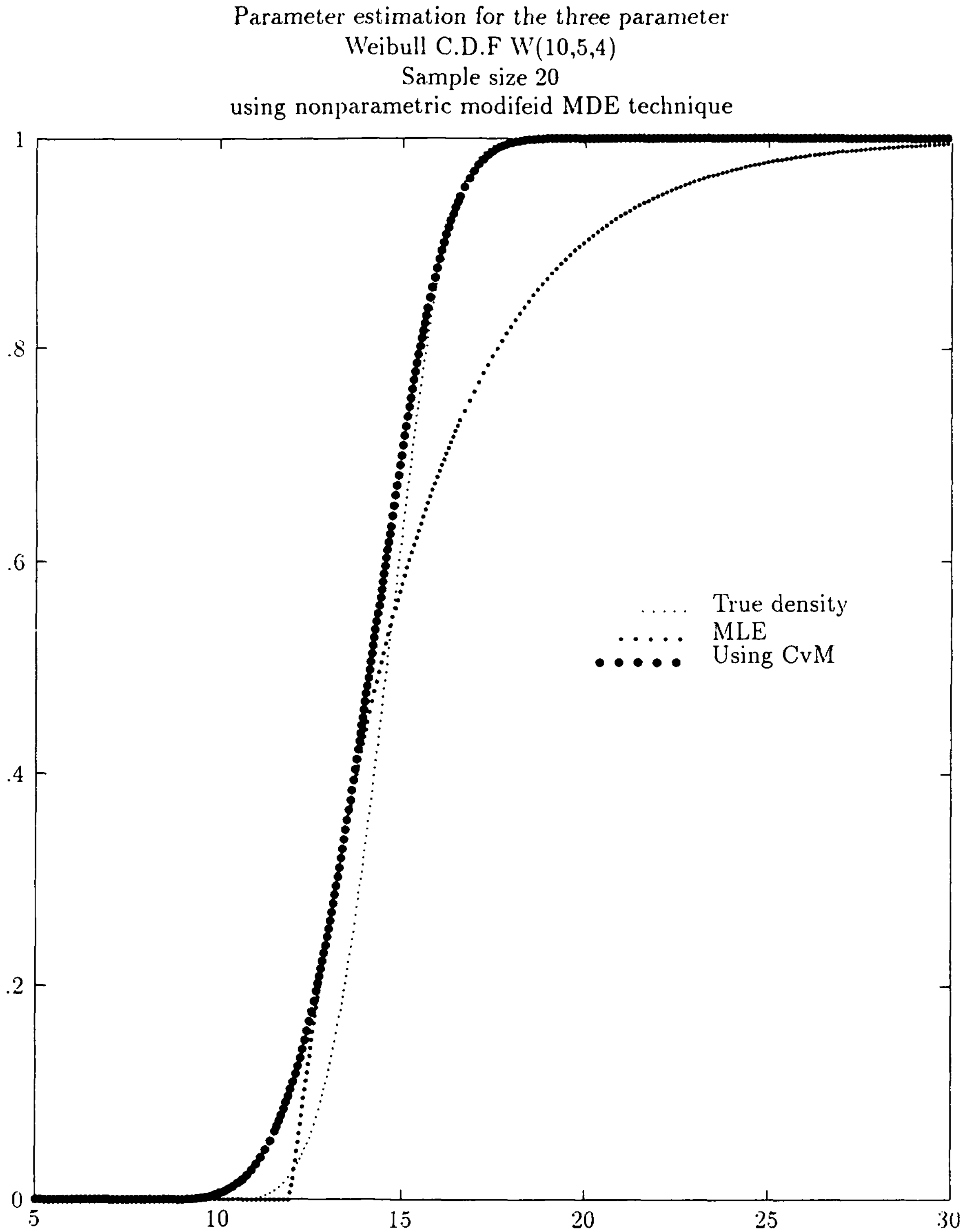

Figure 28. C.D.F. for W(10,5,4) with N=20

Table 15. Weibull Sample (Shape = 1.0 and Sample Size = 30)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 1.0

SAMPLE SIZE = 30

Weibull Data Values

10.058190	11.892060	14.149840
10.227110	11.906630	14.439350
10.360260	11.930340	15.096300
10.423800	12.154340	17.355110
10.537260	12.376410	18.124390
10.761180	12.453510	18.503469
11.346590	12.490680	19.405500
11.719840	12.508960	19.906870
11.759740	13.050770	22.591120
11.876000	13.821920	35.215561

	TRUE	MLE	MDCVM
LOCATION	10.0000	10.0572	7.8721
SCALE	5.0000	3.0323	6.8966
SHAPE	1.0000	1.0012	1.3983
ISE		0.2278	0.0528

Function Norm	540.3060
Window Width	2.5974
Optimal CvM	0.0040

Parameter estimation for the three parameter
Weibull density W(10,5,1)
Sample size 30
using nonparametric modifeid MDE technique

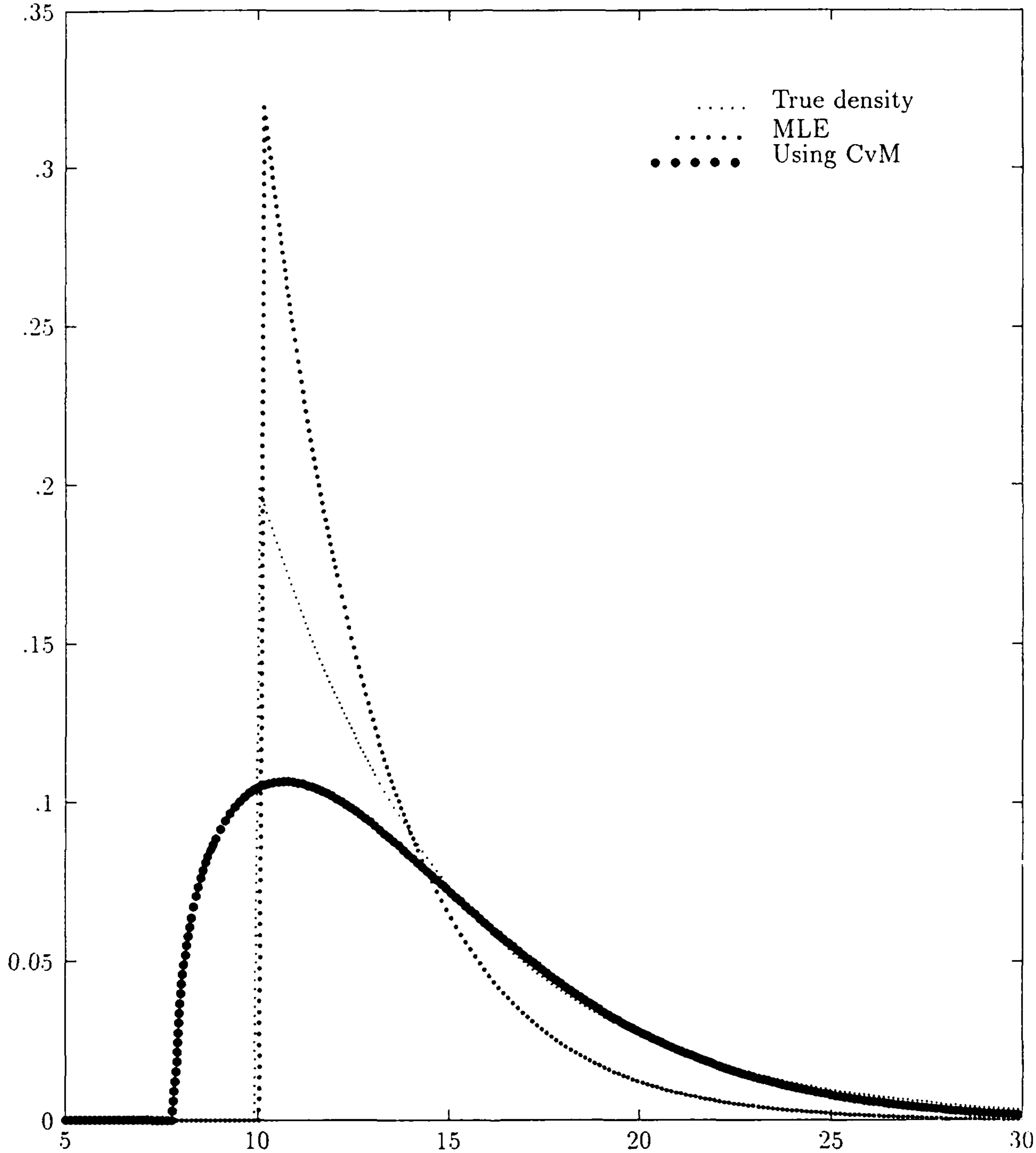

Figure 29. p.d.f for W(10,5,1) with N=30

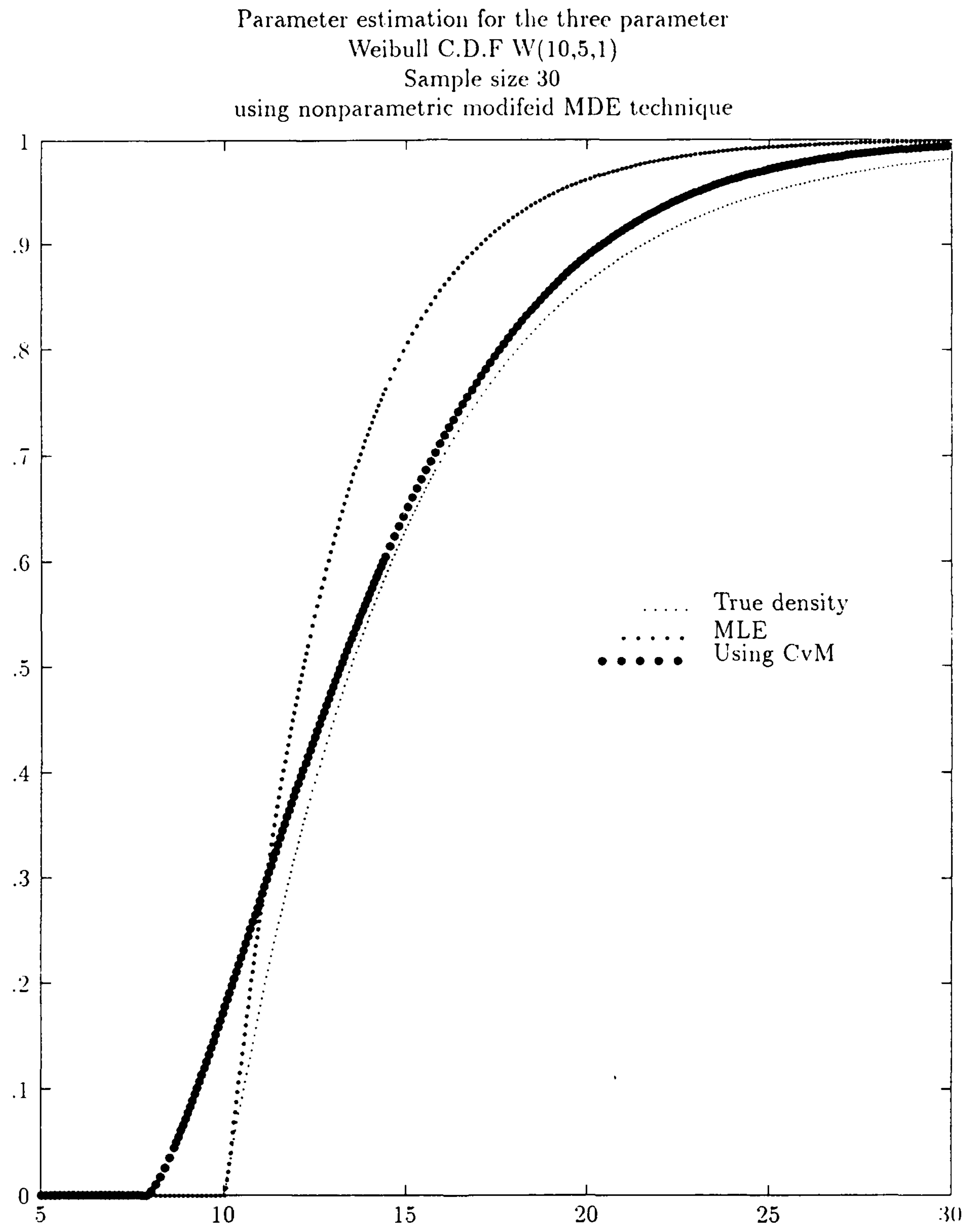

Figure 30. C.D.F. for W(10,5,1) with N=30

Table 16. Weibull Sample (Shape = 2.0 and Sample Size = 30)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 2.0

SAMPLE SIZE = 30

Weibull Data Values

10.539380	13.075760	14.555130
11.065610	13.087580	14.711340
11.342130	13.106710	15.047920
11.455670	13.282030	16.064280
11.638990	13.447030	16.373541
11.950870	13.502510	16.520531
12.594800	13.528930	16.857660
12.932440	13.541860	17.038059
12.966260	13.905620	17.934460
13.062680	14.371460	21.228439

	TRUE	MLE	MDCVM
LOCATION	10.0000	10.5384	9.5878
SCALE	5.0000	2.1066	4.9859
SHAPE	2.0000	1.0111	1.9222
ISE		0.5117	0.0248
Function Norm		18.6620	
Window Width		1.1761	
Optimal CvM		0.0045	

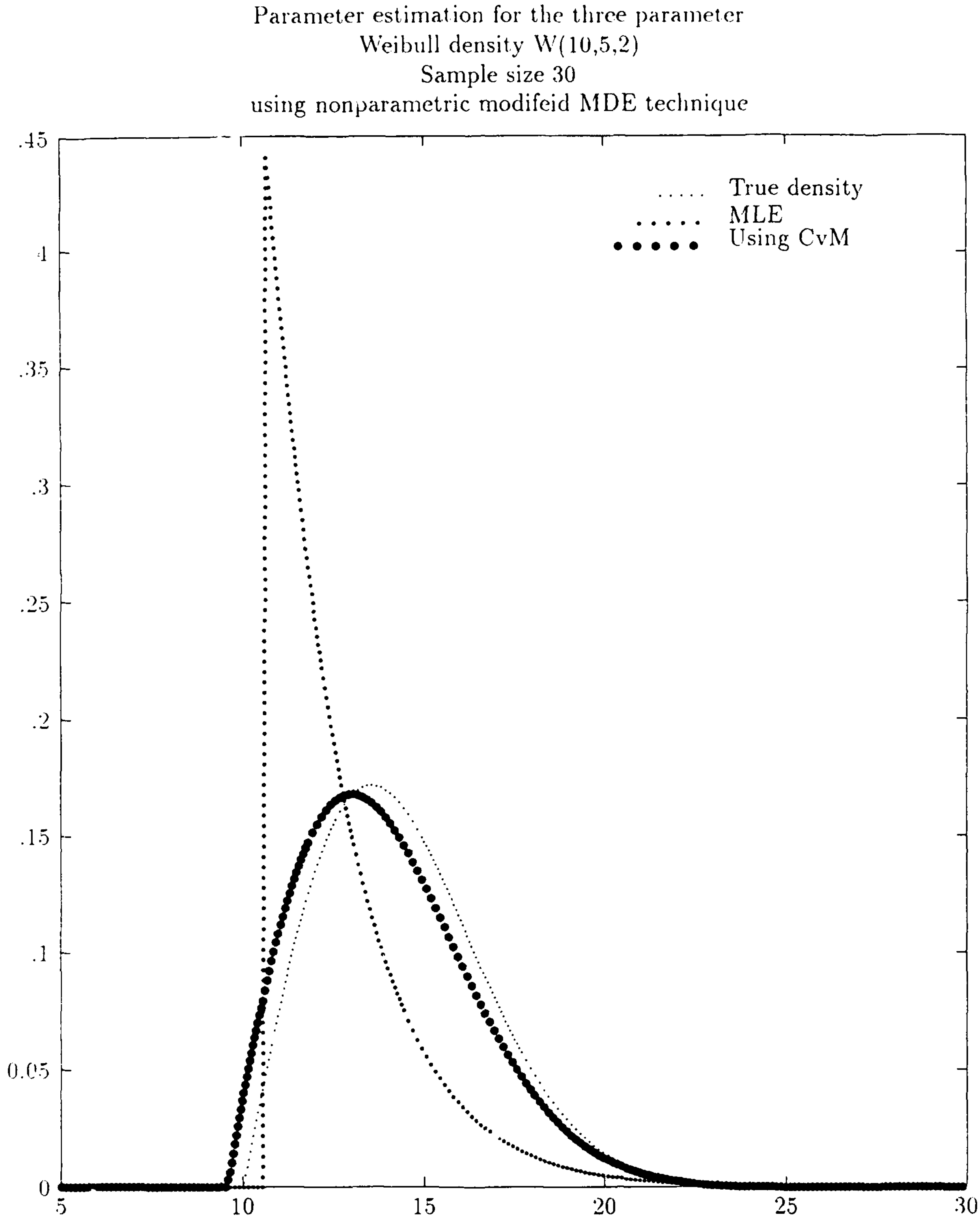

Figure 31. p.d.f for W(10,5,2) with N=30

Parameter estimation for the three parameter
Weibull C.D.F W(10,5,2)
Sample size 30
using nonparametric modifeid MDE technique

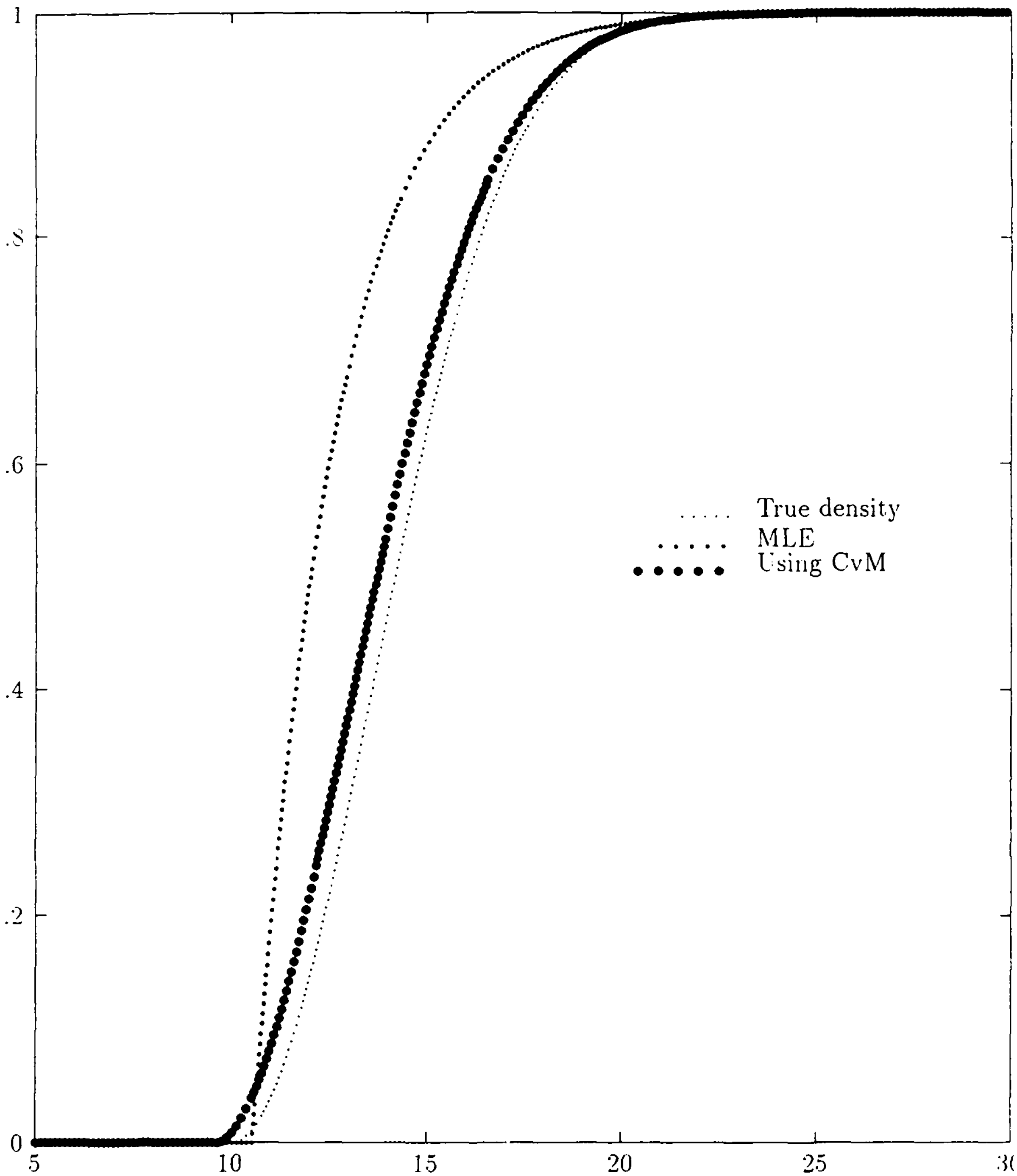

Figure 32. C.D.F. for W(10,5,2) with N=30

Table 17. Weibull Sample (Shape = 3.0 and Sample Size = 30)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 3.0

SAMPLE SIZE = 30

Weibull Data Values

11.133070	13.616520	14.698840
11.783980	13.625780	14.805660
12.080590	13.640750	15.031900
12.196340	13.776460	15.686470
12.377060	13.902000	15.878190
12.669780	13.943750	15.968230
13.228930	13.963560	16.172211
13.503290	13.973240	16.279989
13.530170	14.240820	16.802490
13.606270	14.571660	18.574381

	TRUE	MLE	MDCVM
LOCATION	10.0000	11.1321	11.1331
SCALE	5.0000	2.6142	3.5205
SHAPE	3.0000	1.1138	1.8773
ISE		0.2291	0.0182
Function Norm		10696.4004	
Window Width		0.8241	
Optimal CvM		0.0091	

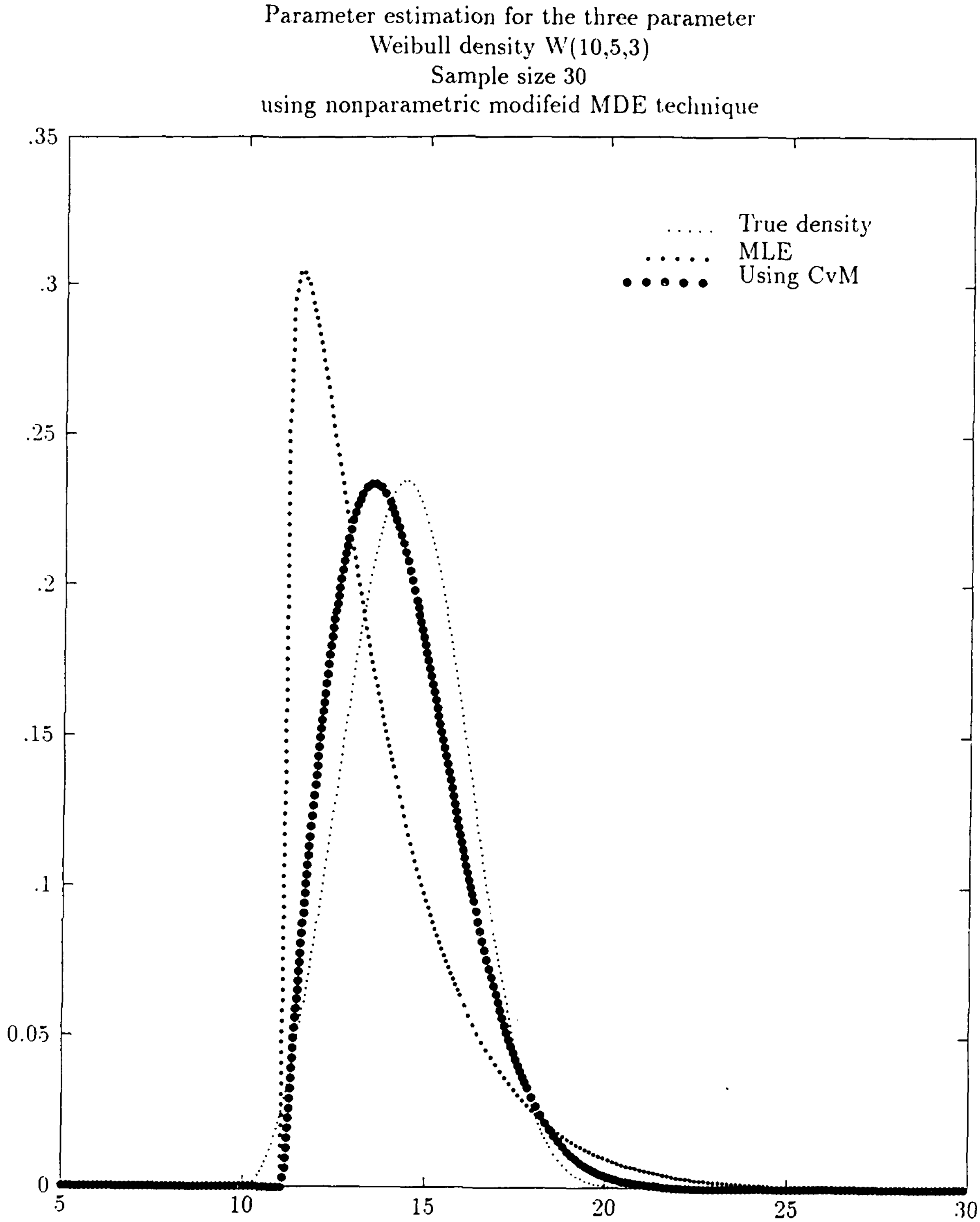

Figure 33. p.d.f for W(10,5,3) with N=30

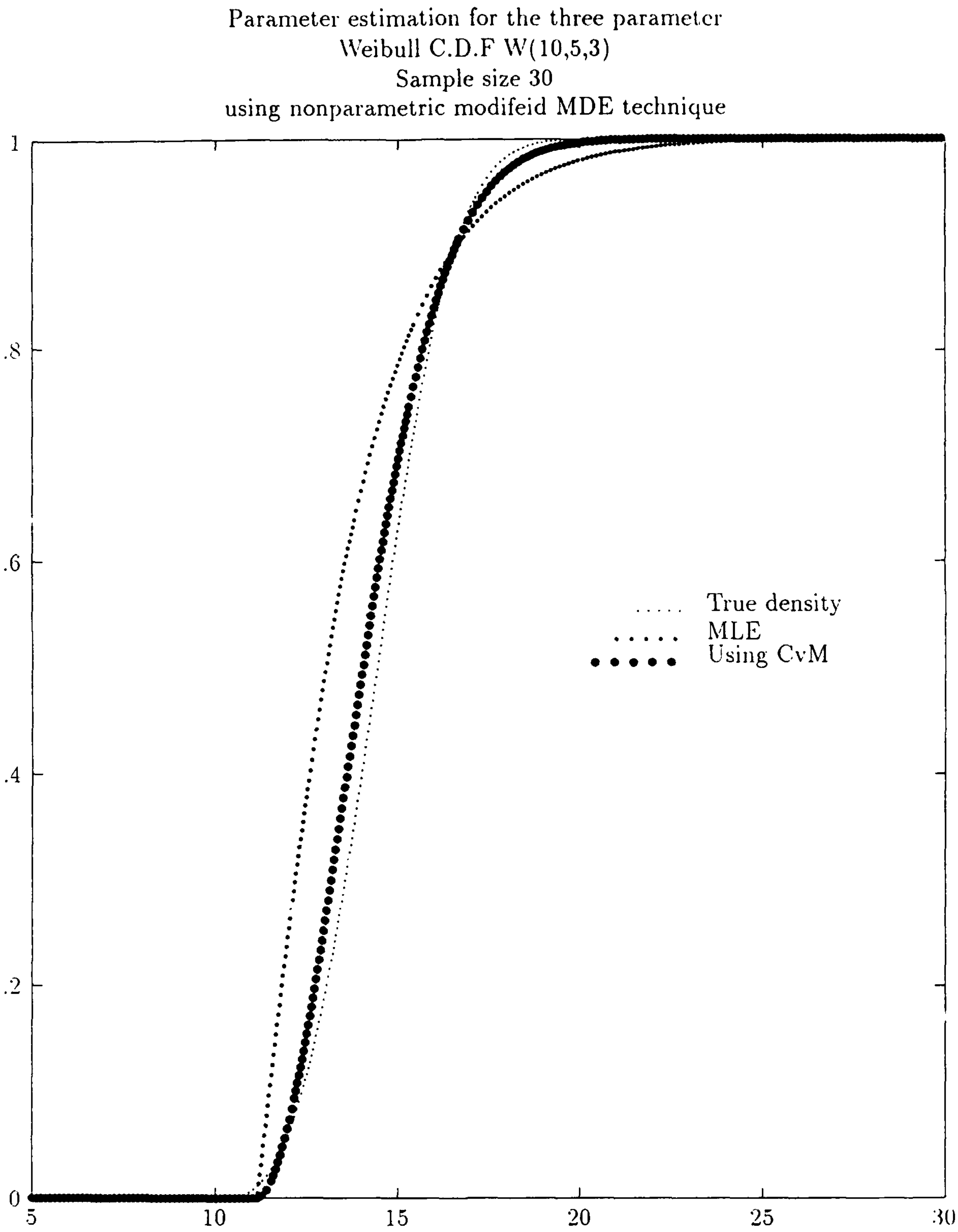

Figure 34. C.D.F. for W(10,5,3) with N=30

Table 18. Weibull Sample (Shape = 4.0 and Sample Size = 30)

TRUE PARAMETERS ARE

Location = 10.0

Scale = 5.0

Shape = 4.0

SAMPLE SIZE = 30

Weibull Data Values

11.963030	13.634270	15.143880
12.212220	14.020930	15.145930
12.269390	14.054020	15.312240
12.331440	14.176460	15.371890
12.840970	14.216850	15.398440
12.860350	14.512920	15.618870
13.062320	14.717380	15.641700
13.130750	14.755560	16.008890
13.219490	14.892670	16.538071
13.540850	14.919950	16.557470

	TRUE	MLE	MDCVM
LOCATION	10.0000	11.9620	7.7954
SCALE	5.0000	3.3463	7.0515
SHAPE	4.0000	1.0086	4.7788
ISE		0.1633	0.0195
Function Norm			90.9906
Window Width			0.6636
Optimal CvM			0.0048

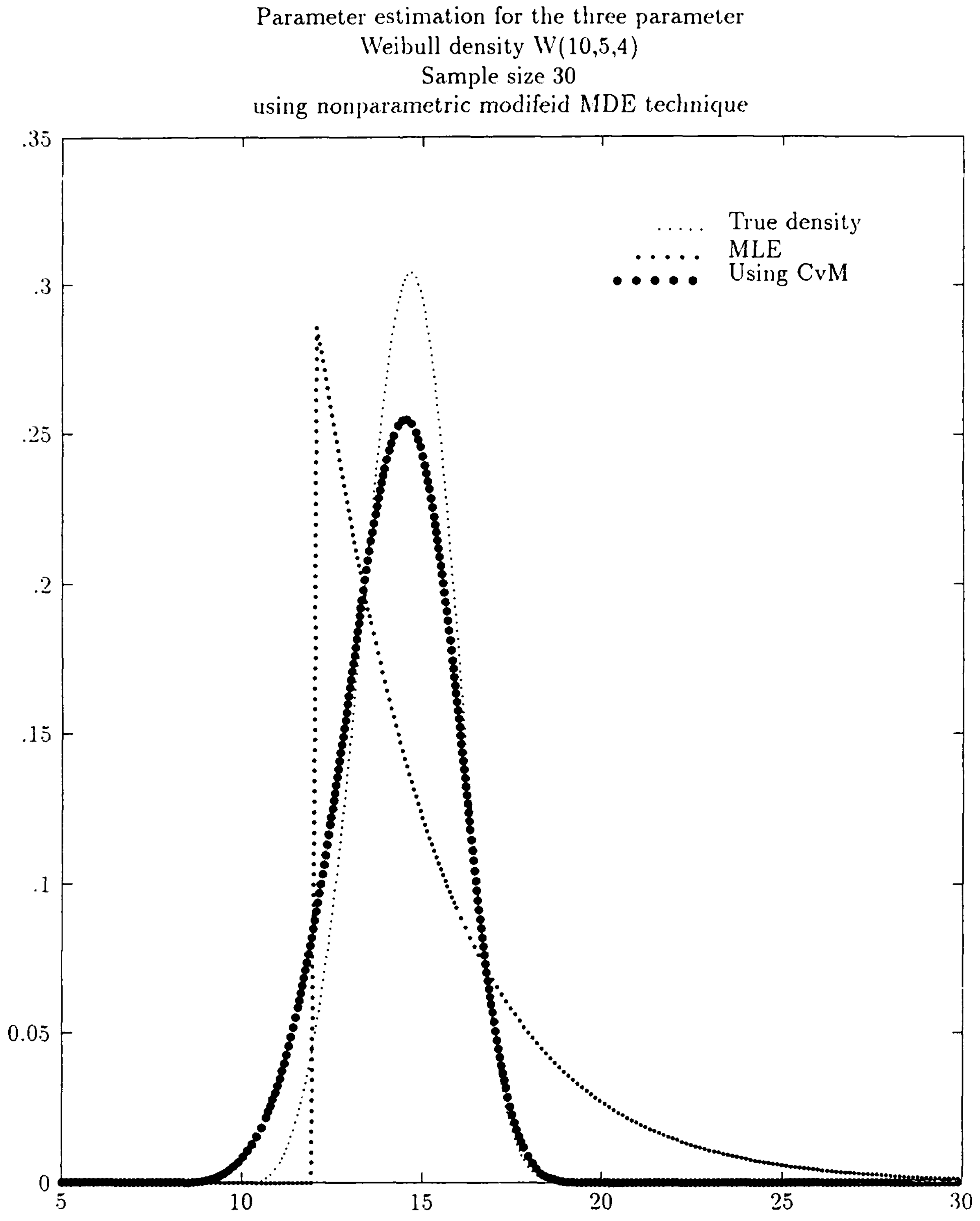

Figure 35. p.d.f for W(10,5,4) with N=30

Parameter estimation for the three parameter
Weibull C.D.F W(10,5,4)
Sample size 30
using nonparametric modifeid MDE technique

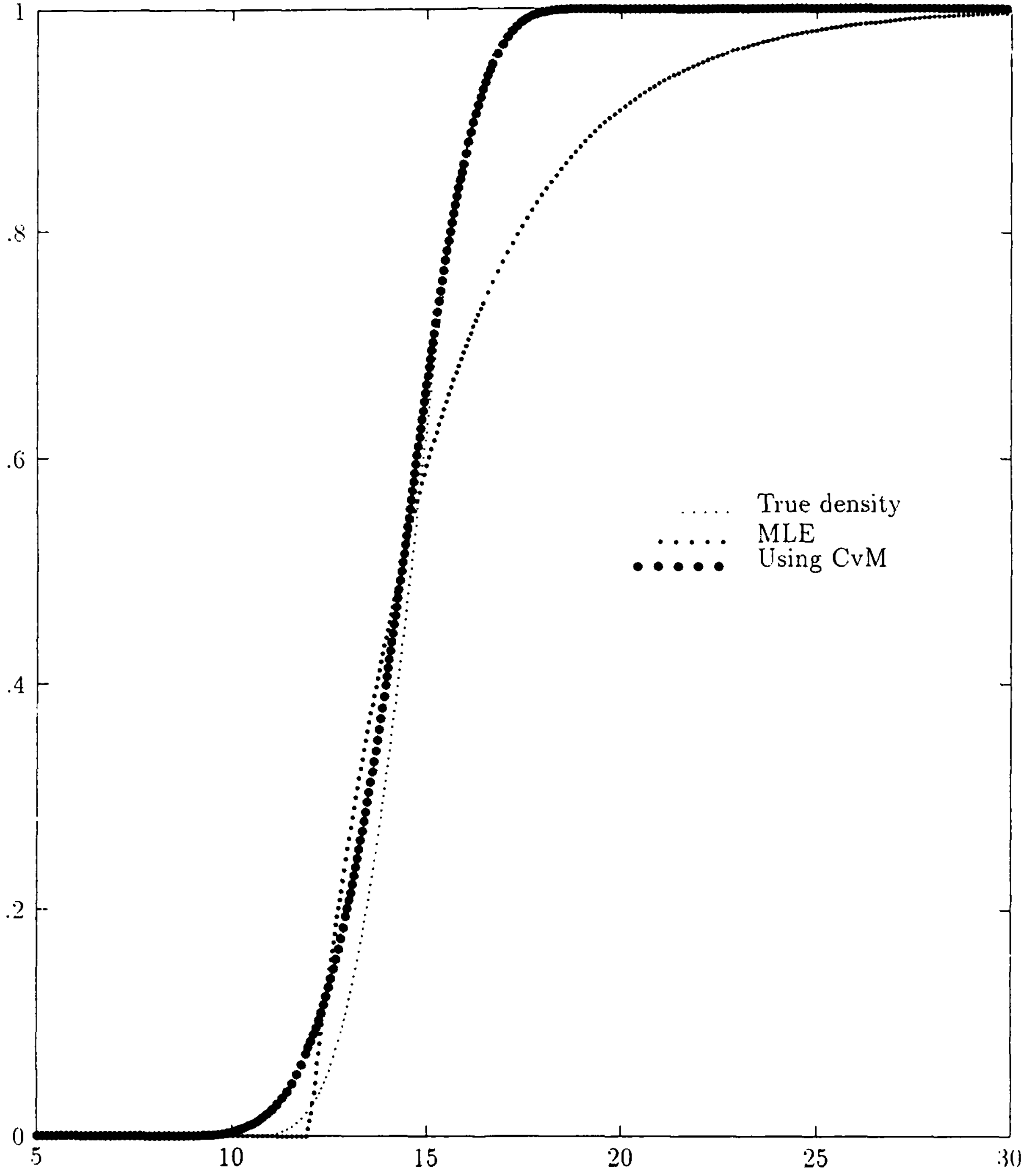

Figure 36. C.D.F for W(10,5,4) with N=30

VII. GOODNESS OF FIT APPLICATION

Introduction

When a sample is drawn from a certain distribution it is hoped that its empirical distribution function (E.D.F) will resemble the population cummulative distribution function (C.D.F). The resemblance, here should take a quantitative meaning. This is done by measuring the closeness or the distance of the E.D.F to the C.D.F. Thus, if $\hat{\hat{F}}_n(x)$ represents the E.D.F and $F_o(x)$ represents the true theoretical C.D.F then many different ways of considering the distances between $\hat{\hat{F}}_n(x)$ and $F_o(x)$ suggest a wide class of goodness of fit statistics. Gini's index of dissimilarity as the integrated absolute difference between both C.D.F's is among these fitting criterion between $\hat{\hat{F}}_n(x)$ and $F_o(x)$. This index is also modified by weighting the integral by $F_o'(x)$. Cramer and von Mises introduced the known Cramer-von Mises statistic (CvM) through weighting the integral of the squared difference between the C.D.F's by $F_o'(x)$. Anderson and Darling introduced the Anderson Darling statistic (AD) by weighting Cramer-von Mises integral by $\frac{1}{F_o(x)[1-F_o(x)]}$. Watson, Kolmogrov and Smirnov, and Kuiper are also other examples of the goodness of fit statistics. Such statistics are surveyed by Stephens (1974). The computational formulae of some of these statistics are given below:

K-S Statistic $\hat{K}$

$$\hat{K} = max \ (D^+ , D^-) \tag{127}$$

where,

$$D^+ = sup(i/n - F_i) \tag{128}$$

$$D^- = sup\left[F_i - \left(\frac{i-1}{n}\right)\right] \quad 1 \leq i \leq n \tag{129}$$

and F_i is F_o at the i^{th} order statistic.

Anderson - Darling $\hat{A}^2$

$$\hat{A}^2 = -n - 1/n \sum_{i=1}^{n} (2i-1) \left[lnF_i + ln\left(1 - F_{n+1-i}\right)\right] \tag{130}$$

Cramer von Mises statistic $\hat{W}^2$

$$\hat{W}^2 = \sum_{i=1}^{n} \left[F_i - \frac{(2i-1)}{2n}\right]^2 + \frac{1}{12n} \tag{131}$$

Kuiper statistic $\hat{V}$

$$\hat{V} = D^+ + D^- \tag{132}$$

The Watson statistic $\hat{U}^2$

$$\hat{U}^2 = \hat{W}^2 - n(F - .5)^2 \tag{133}$$

where

$$\hat{F} = \sum_{i=1}^{n} \frac{F_n}{n} \tag{134}$$

Modified Goodness Of Fit Test

A goodness of fit test based on the E.D.F., where the parameters are estimated is called a modified goodness of fit test.

Basic Characterization

1) If the tables for completely specified null hypothesis are used while the parameters are estimated, this makes the actual α error much smaller and biases the test towards accepting H_0 even without testing.

2) When the parameters are estimated, the null distribution of the test statistic and hence the percentage points will not depend on the location or scale parameter. However, one must use the same estimators as were used in the construction of the tables.

Now, let

$F\left(\frac{x-c}{\theta}\right)$ be a family of C.D.F's with location and scale parameter c, θ respectively, and $F_0\left[\frac{x_i-\hat{c}}{\hat{\theta}}\right]$ the C.D.F when inserting estimators for c, θ under H_0 or simply

denoted as $\hat{F}_i$.

Although the distribution of the test statistic and its percentage points do not depend on c and θ, one should use tables with the same estimators as those used to construct the tables.

A modified K-S goodness of fit test by Monte Carlo simulation for the normal distribution with μ, σ^2 (Lilliefors, 1966) and for the exponential distribution with unknown mean (Lilliefors, 1967) were introduced with a study of the power of the test which showed that the modified K-S test had higher power than χ^2-test for the normal case.

Woodruff et al. (1983) and Bush et al. (1983) derived tables for modified K-S, CvM and AD tests for the Weibull distribution with shape parameter 1 (two parameter negative exponential). Their study showed that the CvM test had the highest power for most of the alternative distributions studied when the null hypotheses was the two parameter negative exponential. They, in addition, studied Weibull with different shape parameters and showed that the AD statistic was the most powerful when the null distribution was Weibull with shape parameter 3.5. A relationship between the critical values and the inverse of the shape parameter was presented for the range of the shape parameters studied.

As for the two parameter Weibull, a BLUE and BLIE (best linear invariant estimator) for the unknown parameters is found in Mann (1968) using the fact that two parameter Weibull is transformed into extreme value by a logarithmic transfor-

mation. She also derived a goodness of fit test for the extreme value distribution of smallest values.

Tables of critical values for the modified K-S, CvM and AD statistics using M.C techniques for the extreme value distribution where the MLE for the parameters is used are derived in a paper by Littelle et al. in 1979.

Tables for the percentage points for the modified K-S, AD and CvM statistics for the gamma distribution are derived in Woodruff et al. (1984).

In addition, similar tables are derived for the critical values for the modified K-S, AD and CvM goodness of fit for the logistic distribution with unknown shape and location parameter using MLE to estimate the parameters (Woodruff et al., 1986).

Porter and Moore derived tables of critical values for the modified K-S, AD and CvM goodness of fit statistics for the Pareto distribution with unknown shape parameter. The powers were shown for eight alternative distribution. In addition they derived a functional relationship between the shape parameters and the critical values of the test.

Yen and Moore derived tables of critical values for the modified AD and CvM goodness of fit statistics for the Laplace distribution. The critical values were tabled for sample sizes n=5(5)50 and significant levels α=.1,.2,.5. The AD test generally yielded higher power than the CvM test.

Harter et al. (1984) modified the definition of the C.D.F at the i^{th} order statis-

tic to obtain a modified K-S test statistic when the probability model is completely specified. They have shown that their proposed test is more powerful than the usual K-S tests for small to moderate sample sizes.

New goodness of fit tests for symmetric alternatives were obtained by Moore for the normal distribution by using a reflection technique in which the data points are reflected about an invariant estimate of the mean and is used to double the sample size. New tables were derived for the K-S, AD and CvM statistics.

A similar work was done by Woodruff et al. for the uniform distribution and by Yen and Moore for the Laplace distribution.

As a final note, a problem arises when a goodness of fit test fails to reject two families of distributions which means that the test does not sufficiently discriminate these two families. Bain used a likelihood ratio test to discriminate normal versus two parameter exponential; normal versus double exponential; normal versus Cauchy or Weibull versus lognormal and extreme value versus normal.

Methodology

Two basic test Statistics are used in this application. This statistics are based on the Cramer von Mises and the Anderson Darling statistics.

(1) Cramer von Mises statistic

The Cramer-von Mises is defined as $W_n^2 = \int_{-\infty}^{\infty} \left[\hat{F}_n(x) - F_o(x)\right]^2 dF_o(x)$. This statistic has the well known classical results that:

- $W_n^{\prime 2}$ is a directed distance.

i.e for any proper distribution function $F_1(x) and F_2(x)$:

$$W_n^2(F_1, F_2) = 0 \Longleftrightarrow F_1(x) = F_2(x) \tag{135}$$

and

$$W_n^2(F_1, F_2) + W_n^2(F_1, F^*) \geq W_n^2(F_1, F^*) \tag{136}$$

- $W_n^{\prime 2}$ is symmetric.

i.e if $F_1(.)$ and $F_2(.)$ are continuous then:

$$W_n^2(F_1, F_2) = W_n^2(F_2, F_1) \tag{137}$$

- Since $0 \leq F(x) \leq 1 \Longrightarrow 0 \leq W^2(F_1, F_2) \leq \frac{1}{3}$.

(2) **Anderson - Darling $\hat{A}^2$**

The AD statistic, considered one of the Cramer-von Mises family, is defined as:

$$Q = \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 \, \Psi(x) \, d\,F(x) \tag{138}$$

where $\Psi(x)$ is some function that weights the square of the difference between both distribution functions. The CvM statistic sets this weight equal to 1. While the AD

statistic uses this weight as the ratio between F(x) and 1-F(x)

$$A^2 = -n - (1/n)\sum_{i=1}^{n}(2i-1)\left[log\left(F(X_{(i)})\right) + log\left\{1 - F(X_{(n+1-i)})\right\}\right] \tag{139}$$

$$\hat{A}^2 = -n - 1/n\sum_{i=1}^{n}(2i-1)\left[ln\hat{F}_i + ln\left(1 - \hat{F}_{n+1-i}\right)\right] \tag{140}$$

In this context a gof test is run using M.C size 1000. The test is based on the AD test statistic where the nonparametric probability is used in place of the EDF. The AD statistic is more sensitive to the distribution tail length by the construction of the weight function above. As for the properties of the E.D.F upon which it seemed natural to use the E.D.F for goodness of fit of $F_0(x)$ (*the theoretical distribution*) is its uniform convergence and almost surely to $F_0(x)$. Subjectively it can be stated that reject if W^2 is large and accept when it is small.

In the application here $F_0(x)$ is assumed to be univariate continuous distribution function. This means that $F_0(X_i)$ will be uniformly distributed between (0,1). The asymptotic behavior of W_n^2 when $F_0(x) = F(x)$ is given by:

$$P\left\{nW_n^2 \leq x\right\} \longrightarrow F_{W_n^2}(x) \tag{141}$$

where

$$F_{W{n^2}}(x) = \frac{1}{\pi\sqrt{x}} \tag{142}$$

The technique used is based upon the idea of using the nonparametric density estimator in place of the E.D.F. Hence the goodness of fit application documents this other new application with complete test elements.

The Technique And The Results

The Monte Carlo procedure for this test was divided to 3 basic stages

Stage I

(1) Determine the Critical Values for the test Statistic at the predetermined significance levels (.01 , .05 (.05) .20).

(2) Compute the value of W^{*2} for each of the 1000 M.C cases as a measure of the distance between the parametric density with the maximum likelihood estimator for the parameters $(\bar{x}, s^2)$ and a nonparametric fit for each sample.

(3)The 1000 M.C samples yields a corresponding sample of size 1000 for W^{*2}.

-Thus, there are two ways to go to find the distribution of W^{*2}:

(a) Use a plotting position.

(b) Fit a continuous nonparametric distribution.

(4) The nonparametric fit is used and the inverse function of the corresponding C.D.F is computed at the different levels of significance.

Stage II

(1) The corresponding power study for the hypotheses is conducted under H_0 and the power is computed.

(2) The test shows powers which were reasonably close to the α-levels

Stage III

The members of the following family of distributions is used as alternative distributions:

-Uniform

-χ^2 with 1 d.f

-χ^2 with 4 d.f

-Exponential

-Cauchy

-D.E

-t-student with 3 d.f

-Logistic distribution

This family of distributions give a variety of shapes and characteristics. The results from this part are shown in the following tables. The tables give the critical values for both cases when the CvM statistic is used and when the AD statistic is used. The tables also show the power of both tests for different sample sizes.

Table 19. Critical Value for the New Suggested Test
for Sample Size = 5 (5) 60
(Using CvM)
(at Significance Levels .2,.15,.1,.05,.01)

N	0.20	0.15	0.10	0.05	.01
5	0.0341	0.0352	0.0364	0.0382	0.0406
10	0.0335	0.0355	0.0387	0.0437	0.0507
15	0.0355	0.0385	0.0418	0.0478	0.0568
20	0.0384	0.0420	0.0455	0.0533	0.0717
25	0.0397	0.0436	0.0495	0.0568	0.0742
30	0.0414	0.0459	0.0508	0.0599	0.0739
35	0.0419	0.0467	0.0520	0.0623	0.0786
40	0.0447	0.0485	0.0554	0.0654	0.0874
45	0.0473	0.0522	0.0590	0.0719	0.0918
50	0.0487	0.0534	0.0600	0.0695	0.0989
55	0.0504	0.0556	0.0637	0.0753	0.0970
60	0.0510	0.0563	0.0639	0.0771	0.0977

Table 20. Power of Tests for Normal Distriution
with Sample Size = 5 (5) 60
(Using CvM)
(at Significance Levels .2,.15,.1,.05,.01)

N	0.20	0.15	0.10	0.05	.01
5	0.2072	0.1667	0.1024	0.0572	0.0121
10	0.2209	0.1559	0.1083	0.0546	0.0105
15	0.2115	0.1575	0.1032	0.0481	0.0105
20	0.1979	0.1485	0.1000	0.0521	0.0103
25	0.1990	0.1504	0.0968	0.0505	0.0099
30	0.2041	0.1455	0.0975	0.0515	0.0101
35	0.1940	0.1489	0.1018	0.0503	0.0096
40	0.2011	0.1502	0.0968	0.0482	0.0100
45	0.1997	0.1392	0.0957	0.0476	0.0102
50	0.1951	0.1427	0.1018	0.0491	0.0098
55	0.1917	0.1505	0.0986	0.0498	0.0096
60	0.1965	0.1477	0.0972	0.0483	0.0097

Table 21. Power of Tests for Normal Distriution with Sample Size = 5

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.2590	0.4480	0.2330	0.3170
.15	0.1990	0.3860	0.1760	0.2530
.10	0.1480	0.3270	0.1420	0.2040
.05	0.0700	0.2510	0.0630	0.1420
.01	0.0070	0.1170	0.0170	0.0380

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.4060	0.1630	0.2350	0.1640
.15	0.3680	0.1190	0.1900	0.1220
.10	0.3280	0.0850	0.1450	0.0840
.05	0.2550	0.0490	0.0930	0.0430
.01	0.1640	0.0090	0.0320	0.0050

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 22. Power of Tests for Normal Distriution with Sample Size = 10

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.5310	0.8090	0.4520	0.6420
.15	0.4660	0.7660	0.3950	0.5650
.10	0.3320	0.6640	0.2810	0.4450
.05	0.2050	0.5000	0.1630	0.2920
.01	0.0700	0.2750	0.0580	0.1370

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.5280	0.2320	0.2540	0.2110
.15	0.5030	0.1930	0.2120	0.1710
.10	0.4560	0.1310	0.1460	0.1020
.05	0.3790	0.0630	0.1000	0.0470
.01	0.2790	0.0250	0.0450	0.0150

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 23. Power of Tests for Normal Distriution with Sample Size = 15

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.6600	0.9370	0.5550	0.7940
.15	0.5810	0.9050	0.4750	0.7400
.10	0.5040	0.8540	0.3800	0.6590
.05	0.3500	0.7610	0.2830	0.5180
.01	0.1590	0.5730	0.1260	0.3030

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.6390	0.2040	0.2560	0.2040
.15	0.6110	0.1640	0.2060	0.1500
.10	0.5810	0.1230	0.1590	0.1110
.05	0.5320	0.0770	0.1060	0.0600
.01	0.4540	0.0360	0.0560	0.0190

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 24. Power of Tests for Normal Distriution with Sample Size = 20

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.7530	0.9800	0.6740	0.8580
.15	0.6670	0.9690	0.5910	0.8130
.10	0.5820	0.9550	0.4990	0.7580
.05	0.3930	0.8890	0.3500	0.6300
.01	0.1070	0.6060	0.1240	0.3100

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.6610	0.1930	0.2550	0.1760
.15	0.6410	0.1560	0.2140	0.1320
.10	0.6160	0.1210	0.1850	0.1100
.05	0.5690	0.0800	0.1350	0.0510
.01	0.4670	0.0190	0.0620	0.0060

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 25. Power of Tests for Normal Distriution with Sample Size = 25

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.8360	0.9940	0.7570	0.9250
.15	0.7680	0.9900	0.7100	0.8950
.10	0.6360	0.9790	0.6090	0.8350
.05	0.5050	0.9520	0.4780	0.7520
.01	0.1910	0.8070	0.2340	0.4980

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.7410	0.1790	0.2530	0.1610
.15	0.7130	0.1330	0.2200	0.1180
.10	0.6840	0.0940	0.1700	0.0780
.05	0.6420	0.0650	0.1270	0.0400
.01	0.5640	0.0260	0.0740	0.0040

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 26. Power of Tests for Normal Distriution with Sample Size = 30

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.8720	0.9980	0.8140	0.9690
.15	0.8210	0.9950	0.7350	0.9530
.10	0.7500	0.9880	0.6600	0.9290
.05	0.6010	0.9740	0.5370	0.8540
.01	0.3470	0.9200	0.3340	0.6900

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.8080	0.1680	0.2690	0.1430
.15	0.7740	0.1390	0.2320	0.0990
.10	0.7450	0.1000	0.1980	0.0680
.05	0.6760	0.0690	0.1550	0.0330
.01	0.6080	0.0360	0.1120	0.0100

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 27. Power of Tests for Normal Distriution with Sample Size = 35

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.9170	1.0000	0.8740	0.9800
.15	0.8770	0.9990	0.8310	0.9690
.10	0.8280	0.9980	0.7780	0.9470
.05	0.6830	0.9900	0.6440	0.8750
.01	0.4080	0.9620	0.4230	0.7420

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.8840	0.1720	0.3040	0.1450
.15	0.8610	0.1380	0.2630	0.1020
.10	0.8350	0.1100	0.2380	0.0750
.05	0.7720	0.0780	0.1860	0.0370
.01	0.6830	0.0360	0.1220	0.0080

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 28. Power of Tests for Normal Distriution with Sample Size = 40

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.9370	0.9990	0.9000	0.9900
.15	0.9180	0.9990	0.8680	0.9830
.10	0.8560	0.9980	0.7830	0.9710
.05	0.7380	0.9960	0.6740	0.9320
.01	0.3820	0.9750	0.4380	0.8260

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9390	0.1710	0.3110	0.1290
.15	0.9200	0.1480	0.2800	0.0990
.10	0.8910	0.1090	0.2340	0.0640
.05	0.8420	0.0710	0.1860	0.0360
.01	0.7380	0.0280	0.1250	0.0070

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 29. Power of Tests for Normal Distriution with Sample Size = 45

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.9530	1.0000	0.9040	0.9930
.15	0.9320	1.0000	0.8720	0.9890
.10	0.8910	1.0000	0.8130	0.9780
.05	0.7390	0.9990	0.7010	0.9520
.01	0.4480	0.9900	0.5110	0.8710

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9410	0.1730	0.3100	0.1250
.15	0.9290	0.1460	0.2770	0.0870
.10	0.9120	0.0980	0.2370	0.0620
.05	0.8770	0.0670	0.1800	0.0270
.01	0.7990	0.0290	0.1250	0.0080

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 30. Power of Tests for Normal Distriution with Sample Size = 50

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.9630	1.0000	0.9410	0.9990
.15	0.9440	1.0000	0.9210	0.9990
.10	0.9170	1.0000	0.8890	0.9960
.05	0.8360	0.9990	0.8170	0.9860
.01	0.4620	0.9910	0.5450	0.9040

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9610	0.1920	0.3260	0.1240
.15	0.9540	0.1610	0.2970	0.0950
.10	0.9440	0.1250	0.2520	0.0610
.05	0.9250	0.0780	0.1990	0.0300
.01	0.8330	0.0190	0.1310	0.0060

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 31. Power of Tests for Normal Distriution with Sample Size = 55

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.9810	1.0000	0.9600	0.9990
.15	0.9600	1.0000	0.9490	0.9980
.10	0.9190	1.0000	0.9110	0.9940
.05	0.8430	1.0000	0.8400	0.9870
.01	0.5910	1.0000	0.6620	0.9580

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9800	0.2030	0.3420	0.1050
.15	0.9730	0.1710	0.3180	0.0820
.10	0.9600	0.1190	0.2760	0.0440
.05	0.9440	0.0690	0.2340	0.0200
.01	0.8970	0.0300	0.1680	0.0040

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 32. Power of Tests for Normal Distriution with Sample Size = 60

(Using CvM)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.9830	1.0000	0.9810	1.0000
.15	0.9720	1.0000	0.9770	0.9990
.10	0.9480	1.0000	0.9550	0.9980
.05	0.8780	1.0000	0.8890	0.9960
.01	0.6750	0.9990	0.7290	0.9740

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9850	0.2180	0.3590	0.1120
.15	0.9810	0.1700	0.3200	0.0760
.10	0.9730	0.1360	0.2840	0.0520
.05	0.9600	0.0770	0.2280	0.0200
.01	0.9290	0.0330	0.1750	0.0080

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 33. Critical Value for the New Suggested Test
for Sample Size = 5 (5) 60
(Using AD)

(at Significance Levels .2,.15,.1,.05,.01)

N	0.20	0.15	0.10	0.05	.01
5	1.1629	1.2454	1.4128	1.6494	2.1284
10	1.5496	1.6466	1.7901	2.1274	2.8964
15	1.9488	2.0551	2.1764	2.4629	3.2314
20	2.2129	2.3563	2.5602	2.8515	3.6844
25	2.4551	2.5653	2.7135	3.0238	3.6552
30	2.6596	2.7707	2.9593	3.2455	4.1849
35	2.8755	2.9862	3.1885	3.4801	4.2493
40	3.0863	3.2155	3.3152	3.7069	4.3719
45	3.2613	3.4019	3.5699	3.8462	4.6924
50	3.4458	3.5522	3.7423	4.0242	4.6589
55	3.6080	3.7104	3.9092	4.1381	4.8513
60	3.7494	3.8857	4.0693	4.3398	4.9610

Table 34. Power of Tests for Normal Distriution
with Sample Size = 5 (5) 60
(Using AD)
(at Significance Levels .2,.15,.1,.05,.01)

N	0.20	0.15	0.10	0.05	.01
5	0.1640	0.1170	0.0600	0.0270	0.0050
10	0.2140	0.1700	0.1100	0.0520	0.0130
15	0.1950	0.1580	0.1140	0.0660	0.0090
20	0.1780	0.1140	0.0630	0.0290	0.0030
25	0.1760	0.1400	0.1030	0.0490	0.0140
30	0.1850	0.1360	0.0830	0.0400	0.0080
35	0.1910	0.1490	0.0870	0.0440	0.0040
40	0.1820	0.1340	0.1070	0.0340	0.0050
45	0.1680	0.1110	0.0660	0.0310	0.0002
50	0.1770	0.1280	0.0780	0.0460	0.0050
55	0.1870	0.1420	0.0960	0.0510	0.0100
60	0.1850	0.1280	0.0900	0.0420	0.0100

Table 35. Power of Tests for Normal Distriution with Sample Size = 5
(Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.1590	0.6210	0.3760	0.4950
.15	0.1140	0.5810	0.3190	0.4460
.10	0.0540	0.4960	0.2220	0.3400
.05	0.0160	0.3710	0.1380	0.2280
.01	0.0050	0.2160	0.0310	0.1000

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.4590	0.1860	0.2740	0.1510
.15	0.4210	0.1320	0.2160	0.0980
.10	0.3370	0.0720	0.1440	0.0570
.05	0.2330	0.0410	0.0820	0.0220
.01	0.1150	0.0090	0.0390	0.0060

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 36. Power of Tests for Normal Distriution with Sample Size = 10
(Using AD)
(Power from K-S test in brackets and with * when better)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.2170 (.2688)	0.9410* (.7850)	0.6460* (.4138)	0.8120* (.5710)
.15	0.1690 (.2112)	0.9280* (.7366)	0.5920* (.3488)	0.7670* (.5120)
.10	0.1090 (.1420)	0.8970* (.6608)	0.5320* (.2716)	0.7180* (.4318)
.05	0.0510 (.0724)	0.8240* (.5420)	0.3720* (.1806)	0.5790* (.3208)
.01	0.0080 (.0128)	0.6350* (.3430)	0.1700* (.0708)	0.3420* (.1612)

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.6980 (.7306)	0.3200 (.3604)	0.3750* (.3610)	0.2530* (.2486)
.15	0.6570 (.6998)	0.2760 (.3030)	0.3190* (.3066)	0.1860 (.1990)
.10	0.5870 (.6532)	0.2090 (.2376)	0.2540* (.2500)	0.1240 (.1418)
.05	0.4510 (.5884)	0.1060 (.1572)	0.1540 (.1726)	0.0590 (.0874)
.01	0.2830 (.4660)	0.0340 (.0646)	0.0590 (.0838)	0.0150 (.0252)

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 37. Power of Tests for Normal Distriution with Sample Size = 15 (Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.2200	0.9910	0.7400	0.9190
.15	0.1720	0.9890	0.6880	0.8960
.10	0.1250	0.9830	0.6400	0.8650
.05	0.0590	0.9590	0.5250	0.7910
.01	0.0100	0.8780	0.2730	0.5940

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.8500	0.3430	0.3940	0.2240
.15	0.8090	0.2940	0.3430	0.1730
.10	0.7780	0.2320	0.2940	0.1210
.05	0.7000	0.1430	0.2110	0.0640
.01	0.5460	0.0370	0.0880	0.0180

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 38. Power of Tests for Normal Distriution with Sample Size = 20 (Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.2440	0.9980	0.8700	0.9660
.15	0.1570	0.9980	0.8240	0.9500
.10	0.1010	0.9960	0.7590	0.9210
.05	0.0500	0.9890	0.6430	0.8760
.01	0.0080	0.9580	0.3610	0.7110

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9160	0.4010	0.4640	0.2470
.15	0.8900	0.3490	0.3970	0.1780
.10	0.8570	0.2520	0.3280	0.1100
.05	0.7900	0.1550	0.2430	0.0510
.01	0.6410	0.0370	0.1090	0.0120

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 39. Power of Tests for Normal Distriution with Sample Size = 25
(Using AD)
(Power from K-S test in brackets and with * when better)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.2800 (.3704)	1.0000* (.9904)	0.9170* (.6566)	0.9960* (.8914)
.15	0.2060 (.2998)	1.0000* (.9860)	0.8850* (.5974)	0.9910* (.8528)
.10	0.1440 (.2156)	1.0000* (.9738)	0.8570* (.5146)	0.9870* (.7960)
.05	0.0710 (.1172)	0.9990* (.9492)	0.7590* (.3872)	0.9550* (.6882)
.01	0.0220 (.0294)	0.9940* (.8484)	0.5710* (.1932)	0.8690* (.4536)

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9680* (.9559)	0.4710 (.5084)	0.5350* (.5138)	0.2630 (.2670)
.15	0.9600* (.9452)	0.4040 (.4402)	0.4680* (.4596)	0.2080 (.2150)
.10	0.9530* (.9298)	0.3250 (.3618)	0.4080* (.3866)	0.1570* (.1494)
.05	0.9190* (.9000)	0.2150 (.2566)	0.2970 (.3004)	0.0880* (.0876)
.01	0.8210* (.8385)	0.0850 (.1196)	0.1620 (.1700)	0.0200 (.0244)

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 40. Power of Tests for Normal Distriution with Sample Size = 30
(Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.3280	1.0000	0.9410	0.9970
.15	0.2340	1.0000	0.9230	0.9970
.10	0.1570	1.0000	0.8790	0.9940
.05	0.0750	1.0000	0.8180	0.9890
.01	0.0130	0.9960	0.5540	0.9130

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9840	0.5210	0.5960	0.2740
.15	0.9820	0.4560	0.5440	0.2200
.10	0.9590	0.3530	0.4570	0.1540
.05	0.9420	0.2390	0.3390	0.0880
.01	0.8370	0.0650	0.1600	0.0120

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 41. Power of Tests for Normal Distriution with Sample Size = 35
(Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.3340	1.0000	0.9670	1.0000
.15	0.2650	1.0000	0.9550	0.9990
.10	0.1630	1.0000	0.9270	0.9960
.05	0.0900	1.0000	0.8740	0.9910
.01	0.0140	0.9990	0.7180	0.9510

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9920	0.5640	0.6460	0.2810
.15	0.9920	0.4950	0.5960	0.2310
.10	0.9840	0.3910	0.4990	0.1450
.05	0.9690	0.2600	0.3960	0.0870
.01	0.9170	0.0810	0.2190	0.0170

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 42. Power of Tests for Normal Distriution with Sample Size = 40
(Using AD)
(Power from K-S test in brackets and with * when better)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.3770 (.5284)	1.0000* (.9896)	0.9880* (.8340)	1.0000* (.9828)
.15	0.2850 (.4482)	1.0000* (.9844)	0.9740* (.7910)	1.0000* (.9752)
.10	0.2290 (.3424)	1.0000* (.9726)	0.9680* (.7248)	1.0000* (.9556)
.05	0.0890 (.1978)	1.0000* (.9490)	0.9050* (.6036)	0.9990* (.9074)
.01	0.0250 (.0454)	1.0000* (.8570)	0.7570* (.3548)	0.9900* (.7204)

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9970* (.9918)	0.6110 (.6376)	0.6650* (.6324)	0.2810 (.2958)
.15	0.9960* (.9888)	0.5300 (.5858)	0.5990* (.5892)	0.2180 (.2450)
.10	0.9950* (.9862)	0.4630 (.5114)	0.5590* (.5132)	0.1810* (.1798)
.05	0.9840* (.9766)	0.2870 (.3852)	0.4200* (.4140)	0.0850 (.1044)
.01	0.9560* (.9498)	0.1090 (.1820)	0.2610* (.2482)	0.0250 (.0312)

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 43. Power of Tests for Normal Distriution with Sample Size = 45
(Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.4360	1.0000	0.9890	1.0000
.15	0.3440	1.0000	0.9810	1.0000
.10	0.2280	1.0000	0.9680	0.9990
.05	0.1410	1.0000	0.9370	0.9990
.01	0.0220	1.0000	0.8100	0.9920

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9990	0.6590	0.7130	0.3210
.15	0.9970	0.5640	0.6540	0.2410
.10	0.9960	0.4730	0.5770	0.1760
.05	0.9900	0.3480	0.4500	0.0900
.01	0.9630	0.1030	0.2350	0.0180

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 44. Power of Tests for Normal Distriution with Sample Size = 50
(Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.4630	1.0000	0.9900	1.0000
.15	0.3860	1.0000	0.9850	1.0000
.10	0.2800	1.0000	0.9720	1.0000
.05	0.1510	1.0000	0.9580	1.0000
.01	0.0390	1.0000	0.8960	0.9980

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	0.9990	0.7140	0.7270	0.3390
.15	0.9990	0.6510	0.6810	0.2720
.10	0.9990	0.5420	0.5940	0.1900
.05	0.9970	0.3990	0.4890	0.0970
.01	0.9820	0.1710	0.3280	0.0270

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 45. Power of Tests for Normal Distriution with Sample Size = 55 (Using AD)

(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.5140	1.0000	0.9970	1.0000
.15	0.4340	1.0000	0.9970	1.0000
.10	0.3090	1.0000	0.9940	1.0000
.05	0.1930	1.0000	0.9840	1.0000
.01	0.0430	1.0000	0.9280	0.9980

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	1.0000	0.7450	0.7480	0.3510
.15	0.9990	0.6850	0.7080	0.2930
.10	0.9990	0.5720	0.6300	0.2060
.05	0.9990	0.4550	0.5480	0.1210
.01	0.9910	0.1910	0.3620	0.0260

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Table 46. Power of Tests for Normal Distriution with Sample Size = 60
(Using AD)
(Power from K-S test in brackets and with * when better or the same)
(Normal against one of the following :)

Sign. level	Uniform	Chi(1)	Chi(4)	Expon.
.20	0.5900 (.6800)	1.0000* (1.000)	1.0000* (.9348)	1.0000* (.9994)
.15	0.4650 (.6012)	1.0000* (1.000)	0.9990* (.9132)	1.0000* (.9984)
.10	0.3380 (.4918)	1.0000* (1.000)	0.9970* (.8640)	1.0000* (.9960)
.05	0.1980 (.3038)	1.0000* (1.000)	0.9940* (.7648)	1.0000* (.9838)
.01	0.0610 (.0952)	1.0000* (.9998)	0.9640* (.5392)	1.0000* (.9312)

Sign. level	Cauchy	D.E	t(3)	Logistic
.20	1.0000* (.9994)	0.7710* (.7536)	0.7810* (.7512)	0.3780* (.3306)
.15	1.0000* (.9990)	0.7010 (.7036)	0.7400* (.7024)	0.2940* (.2736)
.10	1.0000* (.9986)	0.6130 (.6264)	0.6700* (.6356)	0.2130* (.1990)
.05	1.0000* (.9970)	0.4660 (.4816)	0.5780* (.5282)	0.1090 (.1130)
.01	0.9940* (.9926)	0.2290 (.2664)	0.3800* (.3632)	0.0270 (.0342)

Chi(k) = Chi square with k d.f

Expon = Negative Exponential

D.E = Double exponential

t(3) = t-distribution with 3 d.f

Thus, this application defines a new modified goodness of fit test based on the nonparametric kernel density estimator. Both the CvM and AD statistics are used. The critical values are derived by Monte Carlo experiment. Then the power of the test for the case of the CvM and the AD statistics is obtained when the underlying distribution is normal. This power shows a value which is close to the significance level. The test is then performed against each of the eight different alternatives. The power for the different distributions using the CvM statistic shows an incresing power with sample size. The test discriminates all other distributions with high powers, however it does not do as well for the double exponential and the logistic distribution. The modified test using the AD statistic gives better power than the test based on the CvM statistic for the different alternatives except for the uniform distribution due to the fact that the AD statistic is more sensitive to the tails of the distributions than the CvM. The results from the power of the test using AD statistic are compared to those of the classical K-S test for sample sizes 10, 25, 40, and 60. The power from the new modified test using the AD statistic shows an improvement over the classical K-S test in all cases except for the uniform distribution.

VIII. Adaptive Nonparametric Kernel Density Estimation Application

Introduction

In this chapter an "adaptive" approach for the density estimation is introduced. This approach is based on a given criteria according to which a suitable or near optimal adaptive choice of the window width is to be used for the kernel fit.

The general strategy for this application is to generate different samples from various distributions. For each sample a criteria to classify or discriminate the parent distribution from which the sample is drawn is computed. Based on the criteria, a suitable choice of the window width for each case is found. The chosen h value is considered an adaptive choice in this case since it varies with the computed sample criteria. As the adaptive choice for the h parameter is found a nonparametric kernel estimator for the underlying density will be estimated.

Percentile Ratios

For the development of this application, a discriminant was needed. The kurtosis, Hogg's Q statistic, and the percentile ratios are examples of such discriminants that could be used. Since both the kurtosis and the Q statistic average the measure for the upper and lower tail lengths, they are not compatible with the asymmetric distributions. The percentile ratios were chosen to be used as a discriminant since

it measures both tail lengths separately. The upper and lower tail lengths are measured for the distribution by the upper and lower percentile ratios which are defined respectively to be:

$$P_u = \frac{F^{-1}(.975) - F^{-1}(.5)}{F^{-1}(.75) - F^{-1}(.5)} \tag{143}$$

$$P_l = \frac{F^{-1}(.5) - F^{-1}(.025)}{F^{-1}(.5) - F^{-1}(.25)} \tag{144}$$

where

$F^{-1}(\alpha)$ .... represents the α percentile of the distribution.

The population percentile ratios for some distributions with scale parameter zero and shape parameter 1 is given in the following table

Table 47. Values of Percentile ratios for Different Distributions

Distribution	P_l	P_u
Uniform	1.900	1.900
Logistic	3.343	3.343
Exponential	1.647	4.322
Double Exponential	4.322	4.322
Cauchy	12.706	12.706
Normal	2.904	2.904
Beta(1/2,1/2)	1.409	1.409

The median rank is used to find the α sample percentiles based on a sample

of size 60. This gives the sample percentile ratio as:

$$P_u = \frac{a_5 - a_3}{a_4 - a_3} \tag{145}$$

$$P_l = \frac{a_3 - a_1}{a_3 - a_2} \tag{146}$$

where

$$a_1 = .19X_{(1)} + .81X_{(2)} \tag{147}$$

$$a_2 = .6X_{(15)} + .4X_{(16)} \tag{148}$$

$$a_3 = .5X_{(1)} + .5X_{(2)} \tag{149}$$

$$a_4 = .399X_{(1)} + .601X_{(2)} \tag{150}$$

$$a_5 = .81X_{(1)} + .19X_{(2)} \tag{151}$$

where $X_{(l)}$...the l^{th} order statistic

To fit a nonparametric distribution to the given data using the kernel estimation, it is required to find the value of the window width h to be used. A numerically optimal value for various distributions for a sample size 20 is found in Chapter IV. The form used for the h value is:

$$h_{opt} = k\,c\,s\,n^{-\frac{1}{5}} \tag{152}$$

where k...is a constant that varies from one distribution to another.

c...is an adjusting factor for the unbiasedness of the s.

s...is the sample standard deviation.

n...is the sample size.

For this application a sample size n=60 is used for the adaptive method. The form for the optimal h is the same as in chapter IV. The adjusting factor c is 1.0133. The folllowing table gives the values of k for different distributions.

Table 48. Suggested k for the h value

Distribution	k
Uniform	1.0589
Logistic	1.4821
Exponential	.5334
Double Exponential	.8376
Cauchy	.9657
Normal	1.1789

An Adaptive Methodology

A Monte Carlo experiment of size 1000 was performed on a sample size 60 to find the average and the standard deviation of the sample percentile ratios for some distributions. The next table shows the resulting sample average upper and lower percentile ratios with standard deviation given in brackets.

Table 49. Avearge sample percentile ratios

Distribution	P_l	P_u
Uniform	1.9750	1.9513
	(.3937)	*(.4042)*
Exponential	1.7167	4.2838
	(.3172)	*(1.5324)*
Cauchy	83.9780	13.8451
	(634.0062)	*(16.5956)*
Double Exponential	5.2356	4.1618
	(2.0799)	*(1.4824)*
Logistic	3.9424	3.2630
	(1.3424)	*(1.0109)*
Normal	3.2831	2.8688
	(.9372)	*(.7899)*

The adaptive nonparametric density estimation application procedure started by generating 1000 samples each of size 60 from the above distributions. The sample percentile ratios for each sample are then computed. A piecewise linear relation based on the three two tuples (p,k) from the uniform, normal, logistic distributions, where p and k represent the percentile ratios and the constant defined earlier for these distributions respectively is used.

The support is subdivided into three subsets $S_1, S_2\, and\, S_3$ such that $U_{j=1}^{3} S_j =$

Table 50. MISE for the adaptive technique (with standard deviation in brackets)

Distribution	$MISE_{adap}$	$MISE_{III}$
Uniform	.08237	.08267
	(.01891)	*(.02017)*
Logistic	.03244	.03904
	(.01552)	*(.01763)*
Normal	.00870	.00883
	(.00624)	*(.00625)*

$\mathcal{R}$ and such that:

$$S_1 = \{x | x \leq F^{-1}(.25)\} \tag{153}$$

$$S_2 = \{x | F^{-1}(.25) < x < F^{-1}(.75)\} \tag{154}$$

$$S_3 = \{x | x \geq F^{-1}(.75)\} \tag{155}$$

The h value is chosen to vary with each subset of the support. The h is empircally chosen to be a function of the distribution tail length, in the sense of choosing different values of the h for each of the three subsets of the support. This is done by interpolating the piecewise relation for the measured P_l and P_u and finding the corresponding k.

The results from this chapter are shown in table 50. The table gives MISE for this adaptive approach given as $MISE_{adapt}$ and the MISE from chapter III where the estimator for the window width was $sn^{-\frac{1}{5}}$. The table shows that the adaptive

method is doing slightly better in the case of uniform and normal distributions, while for the logistic distribution the method gives 20% improvement in the MISE over that of chapter III. This result depicts that the adaptive technique which is applied for different sample size (60) is working with the values of the constant obtained from the Monte Carlo experiment in chapter IV, and hence could be used in those applications that require no assumption about the distribution form. Hence this chapter gives another tool for applications besides the ones discussed in the earlier chapters.

Appendix A. Generation of random deviates

1. Cauchy Distribution

The probability density function is given by:

$$f(x) = b/\pi \left[(x-a)^2 + b^2\right] \quad a, x \in \mathcal{R}, 0 < b < \infty$$

with

mode(x) = a , median(x) = a

$$\begin{aligned} F(x) &= \int_{-\infty}^{\infty} b/\pi \left[(x-a)^2 + b^2\right] dx \\ &= \frac{1}{\pi} tan^{-1} \left[\frac{(x-a)}{b}\right] + .5 \end{aligned}$$

and the generated deviate will be given by:

$$x = b\,tan\left[\pi\,(u - .5)\right] + a$$

Also, it could be generated using the fact that if (x_1, x_2) are uniformly distributed in a circle centered at the origin then v_1/v_2 will be Cauchy distributed.

2. Logistic Distribution

$$f(x) = exp[-(x-a)/b] / \left[b(1+exp[-(x-a)/b])^2\right]$$

with a C.D.F

$$F(x) = \frac{1}{exp[-(x-a)/b]}$$

with

E(x) = a , V(x) = $\frac{(b\pi)^2}{3}$, mode(x) = a

with variates generated by:

$$x = a - b\,log\left[\left(\frac{1}{u}\right) - 1\right]$$

3. Weibull Distribution

The 3-parameter Weibull density function is given by:

$$f(x) = \frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{\beta-1} exp\left[-\left(\frac{x-\delta}{\theta}\right)^{\beta}\right] \quad , \delta \leq x , \theta, \beta > 0$$

with expected value

$$E(x) = \delta + \theta\Gamma\left(\frac{\beta+1}{\beta}\right)$$

and with variance

$$V(x) = \theta^2 \left[\Gamma\left(\frac{\beta+2}{\beta}\right) - \Gamma^2\left(\frac{\beta+1}{\beta}\right) \right]$$

where Γ denotes the gamma function.

and C.D.F

$$F(x) = 1 - e^{-\left(\frac{x-\delta}{\theta}\right)^{\beta}}$$

and the variates generated by:

$$x = \frac{1}{\theta} exp\left[\frac{ln\left(-ln\left(1-R\right)\right)}{\beta}\right] - \delta$$

BIBLIOGRAPHY

(1) Anderson, G. D. A comparison method for estimating a probability density. Ph.D. dissertation, University of Washington 1969

(2) Bain, L. J. Statistical Analysis of reliability and life testing models (Theory and methods). Marcel Dekker, New York.

(3) Bean, S. J. and Tsokos, C. P. Developments in nonparametric density estimation. Intern. Statist. Rev. 48: 215-235 (1980)

(4) Bickel, P. J.; Doksum, K. A. Mathematical Statistics Holden-Day, Inc. Oakland, California. (1977)

(5) Birge, Lucien On estimating a density using Hellinger distance and some other strange facts. Prob. Theory and Related Fields 71:271-291 (1986)

(6) Bosq, D. Contribution a le theorie de l'estimation fonctionelle. Publ. Instit. Stat. Univ. Paris 19:1-177 (1970)

(7) Boswell, Stephen Nonparametric estimation of the modes of high-dimensional densities. computer Sc. and Stat. :Proceedings of the Sixteenth Symposium:217-225 (1980)

(8) Bowman, A. W. A comparative study of some kernel based nonparametric density estimators . Journal of stat. simul. 21 :313-327 (1985)

(9) Burke, M.; Horvath, L. Density and failure rate estimation in a competing risks model preprint Dept. of Math. and Statist., University of Calgary, Canada.

(1982)

(10) Bush, J. G; Woodruff, B. W.; A. H. Moore et al Modified CvM and AD test for Weibull distribution with unknown location and scale parameters. Comm in Stat. A 12:2463-2476 (1983)

(11) Calitz, Frid An alternative to the Kolmogorov-Smirnov test for goodness of fit. Comm. in Stat. A.16:3519- 3534 (1987)

(12) Cencov, N. N. Evaluation of an unknown distribution density from observations. Soviet. Math. 3:1559-1562 (1962)

(13) Cheng, P. E. A nearest neighbor hazard rate estimator for randomly censored data. Comm. in Stat. B 16:613- 625 (1987)

(14) Cheng, P. E. Hazard rate estimation under the simple proportional hazards model Bulletin of the Institute of Math., Academia Sincia 15:153-162 (1987)

(15) ——— ; Lin, G. W. Maximum likelihood estimation of a survival function under the Kozoil-Green proportional hazard model. Stat. and Prob. Letters 5:75–80 (1987)

(16) D'Agostino, R. B. ; Stephens, M. (Ed); King, Terry (Rev) Review of goodness of fit techniques. Technometrics 29:493-493 (1987)

(17) D'Agostino, R. B. ; Stephens, M. (Ed)Goodness of fit techniques. Marcel Dekker, Inc. New York and Basel (1986)

(18) De Montricher, G. F.; Tapia, R. A. and Thompson, J. R. Nonparametric

maximum likelihood estimation of probability densities by penalty function methods. Ann. Statist. 3:1329-1348 (1975)

(19) Delince, J. Robust density estimation through distance measurements. Ecology, Duke Univ Press, College Station, Box 6697, Durham, NC 27708 67:1576-1581 (1986)

(20)Dennis J. E. JR. ;Schnabel, R. B.Numerical Methods for Unconstrained optimization and Nonlinear equations. Prentice Hall, Inc., Englewood Cliff, N.J. 07632(1982)

(21) Devroye, Luc A course in density estimation. Birkhauser Boston, Inc., Boston, Mass (1987)

(22) ——; Penard, C. S. The strong uniform convergence of multivariate kernel estimates. the Canadian J. of Statistics 14:211-219 1986

(23) ——; Machell, Fred Data structures in kernel density estimation. IEEE Transactions on Pattern Analysis and Machines Intelligence 7:360-366 (1985)

(24) Diggle, Peter; Hall, Peter The selection of terms in an orthogonal series density estimator. JASA 81:230- 233 (1986)

(25) Dodge, Yadolah Some difficulties involving nonparametric estimation of a density function. Journal of Official Stat. 2:193-202 (1986)

(26) Edgeman, R. L.; Scott, R. C. Critical value approximation for an E.D.F-based test of the inverse Gaussian density. Comp. Sc. and Stat:proc. of the 19^{th}

Symp. on the Interface:540-542 (1987)

(27) Eeden, Constance van Mean integrated square error of kernel estimators when the density and its derivative are not necessarily continuous. Annals of the Institute of Statistical Mathematics 37:461-472 (1985)

(28) Efroimovich, S. Yu Nonparametric estimation of a density of unknown smoothness. Theory of Prob. and its Applications 30:557-568 (1986)

(29) Evans, J. W.; Johnson, R. A.; Green, D. W. Two and three parameter Weibull goodness-of-fit tests. Joint Stat. Meeting, ASA 148^{th} Annual meeting, Biometric Society Eastern and Western North Amer. Region (1988)

(30) Fletcher, R. Practical Methods of Optimization second edition John Wiley and Sons. (1987)

(31) Fix, E and Hodges, J.L. Discriminatory analysis, nonparametric estimation: consistency properties. Report No. 4, Project no. 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas, 1951

(32) Fryer, M. J. A review of some non-parametric methods of density estimation. J.Inst math. Appl.20:335-354 (1977)

(33) Fuchs, R. P. A non-parametric probability density estimator and some applications. Ph.D. dissertation, WPAF Base, Ohio AFIT 1984

(34) Gajek, Leslaw On improving density estimators which are not bona fide functions. The Annals of Statistics 14:1612-1618 (1986)

(35) Gallagher, M. A. and Moore, A. H. Robust Minimum distance estimation using the three patameter Weibull. (submitted for publication in IEEE)

(36) Gill, Phillip E.; Murray, W; Wright, M. H. Practical Optimization Academic Press. (1981)

(37) Ghorai, J. K. Nonparametric estimation of probability density function. Ph.D. Dissertation Purdue Univ. 1977

(38) Good, U. J.; Gaskins, R.A. Nonparametric roughness penalties for probability densities. Biometrika 58:255-277 (1971)

(39) —— ; —— A nonparametric estimation of probability densities. Virginia J. of Sci 23:171-193 (1971)

(40) Graybill, F. A. Introduction to Matrices with Applications (1969) in Statistics. Wadsworth Publishing Company, Inc. Belmont, California

(41) Greblicki, Woldzimierz; Pawlak Miroslaw Pointwise consistency of the Hermite series density estimate. Statistics and Prob. Letters 3:65-69 (1985)

(42) Groeneboom, P Estimating a monotone density. Proceedings of the Berkeley Conference in Honor of Jerz Neyman and Jack Kiefer vol 2:539-555 (1985)

(43) Hall, P. Cross-validation and the smoothing of orthogonal series density estimators. J. of Multi. Analysis 21:189-206 (1987)

(44) —— On Kullback-Leibler loss and density estimation. Annals of Stat. 15:1491-1519 (1987)

(45) —— On the use of the compactly supported density estimates in problems of discrimination. J. of Multi. Analysis 23:131-158 (1987)

(46) —— On the rate of convergence of orthogonal series density estimators. Journal of the Royal Statistical Society Series B 48:115-122 (1986)

(47) Hall, P.; Marron, J. S. On the amount of noise inherent in bandwidth selection for a kernel density estimator. Annals of Stat. 15:163-181 (1987)

(48) —— Extent to which least-squares cross-validation minimizes integrated square error in nonparametric density estimation. Prob. Theory 74:567-581 (1987)

(49) Hall, Peter; Watson, G. S.; Cabrera, J. Kernel density estimation with spherical data. Biometrika 74:751-762 (1987)

(50)Harter, H. L.; Khamis, H. T. and Lamb,R.E. Modified K-S tests of goodness of fit. Comm Stat. Simula Comput 13:293-323 (1984)

(51)Harter, H. L.; Moore, A. H. Maximum likelihood estimation of the parameters of gamma and Weibull populations from complete and from censored samples. Technometrics 7 No. 4:639-643 (1965)

(52)Hobbs, J. R.;Moore, A. H. and James, W. Minimum distance estimation of the three parameters gamma distribution. IEEE Transactions on Reliability 33 No.3:237- 240 (1984)

(53)Hobbs, J. R.;Moore, A. H. and Miller, R. M. Minimum distance estimation of the parameters of the three parameter Weibull distribution. IEEE Transactions

on Reliability 34 No. 5:495-496 (1985)

(54) Hosmane, Balakrishna Improved likelihood ratio test for multinomial goodness of fit. Comm. in Stat. A 16: 3185-3198 (1987)

(55) Jinadasa, K. G. Maximum likelihood estimation with additional data. Joint Stat. Meeting, ASA 148th Annual meeting, Biometric Society Eastern and Western North Amer. Region (1988)

(56) Johnson E. G.; Routledge The line transect method: A nonparametric estimator based on shape restrictions. Biometrics 41:669-679 (1985)

(57) Kappenman, R. S. A nonparametric data based univariate density function estimate. Comp. Stat. and Data Analysis 5:1-7 (1987)

(58) Kraft, C. H.; Lepage,Y.; Eeden van, C. Estimation of a symmetric density function. Communications in Statistics A 14:273-288 (1985)

(59) Kullback, Sulomon The Kullback-Leibler distance. Amer. Statistician 41:340-341 (1987)

(60) Kumar, T. K.; Markmann, J. M. Estimation of probability density function : A Monte Carlo comparison of parametric and nonparametric methods.

(61) Littelle, R. D. et al Goodness of fit tests for the two parameter distribution. Comm.Stat. Simul Comp B 8:257-269

(62) Lilliefors,H.W On the Kolmogorov test for the exponential distribution with mean unknown JASA 64:387-389 (1969)

(63) —— On the Kolmogorov test for normality with mean and variance unknown JASA 62:143-147 (1967)

(64) —— The chi square goodness-of-fit test revisited. Joint Stat. Meeting .ASA 148^{th} Annual meeting, Biometric Society Eastern and Western North Amer. Region (1988)

(65) Luong, A.; Thompson, M. E. Minimum-distance methods based on quadratic distances for transforms. Canad. J. of Stat. 15:239-251 (1987)

(66) Man, N. R. Point and interval estimation procedures for the two parameter Weibull and extreme value distributions. Technometrics 10:231-256 (1968)

(67) Marshall, A. W.; Proschan, F. Maximum likelihood estimation for distributions with monotone failure rate. Annals of Math. Stat. 36:69-77 (1965)

(68) McCune, E. D.; McCune, S. K. Modified nonnegative kernel estimator of the failure rate function. Amer. Stat. Assoc., Proc. of Stat. Computing Section:353-354 (1987)

(69) —— On improving convergence rates for nonnegative kernel failure-rate function estimators. Stat. and Prob. Letters 6:71-76 (1987)

(70) Moore, A. H. A modified Kolmogorov-Smirnov test for Weibull distributions with unknown location and scale parameters. IEEE Transactions on Reliability 23:209- 213 (1983)

(71) —— Modified Cramer-von Mises and Anderson- Darling test for Weibull

distribution with unknown location and scale parameters. Communications in Statistics - Theoretical Methods 12(21):2465-2476 (1984)

(72) —— Modified goodness-of-fit tests for gamma distributions with unknown location and scale parameters. IEEE Transactions on Reliability 33:241-245 (1984)

(73) —— A Monte Carlo technique for estimating lower confidence limits on system reliability using pass-fail data. IEEE Transactions on Reliability 32:366-369 (Oct.1983)

(74) —— Minimum-distance estimation of the three parameters of the gamma distribution. IEEE Transactions on Reliability 32:237-240 (Aug.1984)

(75) —— Minimum-distance estimation of the parameters of the 3-parameter Weibull distribution. IEEE Transactions on Reliability 34:495-496 (Dec.1985)

(76) —— A Monte Carlo method for determining confidence bounds on reliability and availability of maintained systems. IEEE Transactions on Reliability 34:497-498 (Dec.1985)

(77) —— Modified goodness-of-fit tests for logistic distribution with unknown location and scale parameters. Comm. in Stat. - Simula. Computa. 15(1):77-83 (1986)

(78) —— Modified goodness-of-fit test for the Laplace distribution. Comm. in Stat. - Simulation 17(1):275- 281 (1988)

(79) —— Extension of Monte Carlo techniques for obtaining system relia-

bilty confidence limits from component test data. Proceedings of national Aerospace electronic conference:459-463 (1965)

(80) —— Robust statistical inference Notes from a short course presented at the Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio (1981)

(81)Moore, A. H.; Ream, J. J. and Woodruff, B. W. A new goodness of fit tests for normality with mean and variance unknown. (submitted for publication)

(82) Moré, J. J.; et al. User Guide for Minipack-1 Argonne National Laboratory.

(83) Nadaraja, E. On nonparametric estimation of density function and regression. Theory Prob. Appl. 10:186-190 (1965)

(84) Ortega, J. M.; Rheinboldt, W. C. Iterative Solution of Nonlinear equations in Several Variables. Academic Press. New York and London. (1970)

(85) Paraska Rao, B. L. S. —underlineNonparametric functional estimation. Academic Press. (1983)

(86) Parr, W. C. and Schucany, W. R. Minimum distance and robust estimation. JASA 75 No 3:616-624 (1980)

(87) Parzen, E. On estimation of a probability function and mode. Annals of Math. Stat. 33:1065-1076 (1962)

(88) Powell, M. J. D. "A hybrid method for nonlinear equations," Numerical methods for Nonlinear Algebraic equations , P. Rabinwitz editor (1970)

(89) Powell, M. J. D. Approximation theory and methods Cambridge University Press, Cambridge (1981)

(90) Porter, J. E.; A. H. Moore et al Modified Kolmogorov, AD and CvM tests for Pareto distribution with unknown location and scale parameters.(submitted for publication)

(91) Ramberg, J. S., et. al. A probability distribution and its use in fitting data. Technometrics 21:201-214 (1979)

(92) Reklaitis, G. V. Engineering optimization. Methods and applications John Wiley Sons (1983)

(93) Revesz, P. Density estimation. Handbook of Statistics vol 4:531-549 (1984)

(94) Rock, N. M. S. NPSTAT: A Fortran-77 program to perform nonparametric variable-by-variable comparisons on two or more independent groups of data. Comp. and Geosc. 12:757-777 (1986)

(95) —— ROBUST: An interactive Fortarn-77 package for exploratory data analysis using parametric, robust and nonparametric location and scale estimates, data transformations, normality tests and outlier assessment. Comp. and Geosc. 13:463-494 (1987)

(96) Schuster, E. F. Note on uniform convergence of density estimates. Annals of Math. Stat. 41:1347-1348 (1970)

(97) Scoff, David W. Choosing smoothing parameters for density estimators.

computer Sc. and Stat. :Proceedings of the Seventeenth Symposium:225-229 (1986)

(98) Sheather, Simon J. An improved criteria for choosing the window width when estimating the density at a point. Comp. Stat. and Data Analysis 4:61-65 (1986)

(99) Silverman, B. W. Density estimation for statistics and data analysis. Chapman and Hall Ltd (1986)

(100) Silverman, B. W.; Young, G. A. The bootstrap: To smooth or not to smooth? Biometrika 74:469-479 (1987)

(101) Stephens, M. E.D.F. statistics for goodness of fit. J. Amer. Statist. 69:730-737 (1974)

(102) Sweeder, J. Nonparametric estimation of distribution and density functions with applications. Ph.D. Dissertation WPAF Base, Ohio AFIT 1982

(103) Tarter, Michael E.; Freeman, William; Hopkins, Alan A Fortran implementation of univariate Fourier series density estimation. Comm. in Stat. B 15:855-870 (1986)

(104) Taylor, Malcolm S ;Thompson, James. R A data based algorithm for the generation of random vectors. Comp. Stat.and Data Analysis A 4:93-101 (1986)

(105) Wahba, Grace Optimal smoothing of density estimates. Annals of Math. Stat.:423-452 (1983)

(106) Wakimoto, Kazumasa et al. Testing the goodness of fit of the multinomial

distribution based on graphical representation. Comp. Stat. and Data Analysis A 5:137- 147 (1987)

(107) Walter, G.; Blum, J. R. Probability density estimation using delta sequences. Annals. of Stat. 7:328-340 (1979)

(108) Wasan, M. T. Parametric Estimation McGraw Hill Book Company.(1970)

(109) Watson, G. S. and Leadbetter, M. R.Hazard analysis I. Biometrika 51:175-184 (1964)

(110) Wegman, E. J. Density estimation. Encyclopedia of Stat Science vol 2:309-315 (1982)

(111) Wolfowitz, J. The minimum distance method. Annals. of Math. Stat. 28:75-88 (1957)

(112) Wolfowitz, J. Estimation by the minimum distance method. Annals. of the Inst of Stat. Math. 5:9-23 (1953)

(113) Woodruff, B. W; A. H. Moore et al A modified K-S test for Weibull distribution with unknown location and scale parameters. IEEE transactions on Reliability:209-213 (1983)

(114) Woodruff, B. W.; A. H. Moore et al Modified goodness fo fit tests for logistic distribution with unknown location and scale parameters.
Stat. Simula Comp 15(1):77-83 (1986)

(115) Woodruff, B. W; A. H. Moore et al A new goodness of fit test for the

uniform with unspecified parameters. (submitted for publication)

(116) Woodruff, B. W.; A. H. Moore et al Modified goodness fo fit tests for gamma distributions with unknown location and scale parameters. IEEE transactions on Reliability 33:241-245 (1984)

(117) Yen, V.C and Moore, A. H Modified goodness fo fit tests for the Laplace distribution.(submitted for publication)

Vita

Col. Ahmed Mohamed M. Sultan [redacted]

He graduated as an electrical engineer in 1975 from the Military Technical College (MTC) of Cairo. Upon the receipt of his degree he served as an electrical engineer for aircraft electrical and special equipment and instrument. Latter on, his interest started to increase in Operations Research (O.R). In 1977 he joined the Institute of Statistical Studies and Research (ISSR) for a two year diploma in O.R. In 1979 he received his diploma degree from the ISSR and worked his graduation project in planning power supply for a new under developed city in Egypt. In 1977 and upon the receipt of his diploma degree he joined the M.S program in O.R for one year in the same school (ISSR). The M.S program in ISSR is a one year of courses and a thesis.

On finishing his first year of courses in ISSR and in Summer of 1981 he was selected on a competitive basis to join a M.S. Program in O.R at the Air Force Institute of Technology (AFIT). In December of 1982 he received his M.S in O.R. He worked his M.S thesis on Robust Multiple Linear Regression, where an extensive Monte Carlo analysis was conducted to determine the performance of robust linear regression techniques with and without outliers.

Upon the receipt of his M.S he worked for the Operations Department of the Egyptian Air Force as an analyst. In addition he worked part time for the M.T.C teaching O.R. and Probability and Statistics.

Latter he was chosen to work as a mathematics instructor in the Egyptian Air Academy and sent for a Ph.D in statistics from AFIT.

Permanent address: 94 Wassef St. Ein Shams
CAIRO , EGYPT

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public REléase Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/DS/ENC/90-1	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Air Force Institue of Technology 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Applications of Non-parametric Density Estimation

12. PERSONAL AUTHOR(S)
Ahmed M.M. Sultan, Lt Col Egyptian Air Force

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Ph.D. Disertation	FROM 10/5/87 TO 2/28/90	February 28, 1990	179

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
12	3	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Distance Estimation, Goodness of Fit Tests Non-Parmetric Density Estimation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The dissertation examines various methods of nonparametric density estimation, and non-parametric kernel estimation in more detail. The consequences of various kernel window width and their effect on the mean integrated square error are examined using Monte Carlo techniques

The mean and the variance of nonparametric density estimator is derived for symmetric kernels with finite mean and finite variance. The results also treat kernels with varying window parameters.

The nonparametric kernel estimate was used to obtain new estimators for the three parameter Weibull distribution using distance estimation and the Cramervon-Mises statistic. Comparison with maximum likelihood estimators using a Monte Carlo sample of size 1000 and various different parameters showed a significant improvement over the maximum liklihood estimators in the mean integrated square error between the estimated distribution and the true distribution.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. ALbert H. MOore	22b. TELEPHONE (Include Area Code) (513)255-3098 / 22c. OFFICE SYMBOL AFIT/ENC

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Serveral new goodness of fit tests are proposed using the nonparametric kernel estimator and the Cramer-von-Mises and the Anderson Darling statistics. Extensive Monte Carlo experiments were performed to obtain the critical values for the test and to study the power of the tests against eight alternative distributions. The tests using the Anderson Darling statistic showed greater power against almost all alternative distributions studied than the K.S. test.

A new nonparametric kernel estimator was introduced by varying the window width in each tail portion of the sample. The method permitted different window width in each tail portion and in the center portion of the sample. The method uses separately the sample percentile ratios as a measure of each tail length. The kernel parameter for the tail sample values is chosen using sample percentile ratios for that tail. The nonparametric kernel estimator results in comparable mean integrated errors with the estimators developed earlier.

Unclassified